

# LARGE DEVIATIONS FOR SYSTEMS WITH NON-UNIFORM STRUCTURE

VAUGHN CLIMENHAGA, DANIEL J. THOMPSON,  
AND KENICHIRO YAMAMOTO

**ABSTRACT.** We use a weak Gibbs property and a weak form of specification to derive level-2 large deviations principles for symbolic systems equipped with a large class of reference measures. This has applications to a broad class of coded systems, including  $\beta$ -shifts,  $S$ -gap shifts, and their factors. Our techniques are suitable for adaptation beyond the symbolic setting.

## 1. INTRODUCTION

We introduce criteria for a symbolic system to satisfy the large deviations principle. These criteria are motivated by the ‘non-uniform’ structure of our main examples –  $\beta$ -shifts,  $S$ -gap shifts, and their factors – but apply more generally. We prove the following main result. (See §2 for precise definitions.)

**Theorem A.** *Let  $(X, \sigma)$  be a shift on a finite alphabet,  $m$  a Borel probability measure on  $X$ , and  $\varphi: X \rightarrow \mathbb{R}$  a continuous function. Let  $\mathcal{L}$  be the language of  $X$ . Suppose there exists a set  $\mathcal{G} \subset \mathcal{L}$  such that*

- [A.1]  $\mathcal{G}$  has  $(W)$ -specification with good concatenations;*
- [A.2]  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ ;*
- [A.3]  $m$  is Gibbs for  $\varphi$  with respect to the collection  $\mathcal{G}$ .*

*Then  $(X, \sigma)$  satisfies a level-2 large deviations principle with reference measure  $m$  and rate function  $q^\varphi: \mathcal{M}(X) \rightarrow [-\infty, 0]$  given by*

$$(1.1) \quad q^\varphi(\mu) = \begin{cases} h(\mu) + \int \varphi d\mu - P(\varphi) & \mu \in \mathcal{M}_\sigma(X), \\ -\infty & \text{otherwise.} \end{cases}$$

Roughly speaking, [A.1] means that words from  $\mathcal{G}$  can be ‘glued together’ with uniformly bounded gaps to obtain another word in  $\mathcal{G}$ . The condition [A.2] means that any word  $w \in \mathcal{L}$  can be transformed

---

*Date:* April 22, 2013.

*2000 Mathematics Subject Classification.* 37A50, 60F10, 37D35, 37D25, 37B10.

*Key words and phrases.* large deviations, specification properties.

D.T. was partially supported by NSF grants DMS-1101576 and DMS-1259311.

into a word in  $\mathcal{G}$  without making too many edits. Finally, **[A.3]** means that  $m$  satisfies an upper Gibbs bound on all cylinders, and a lower Gibbs bound on cylinders corresponding to words in  $\mathcal{G}$ .

**Context of Theorem A.** Large deviations theory originated in probability theory and statistical mechanics, with a long history dating to the 1938 theorem of Cramér (which applies to i.i.d. random variables) and ultimately to a heuristic 1877 observation of Boltzmann [15, 16].

In the context of dynamical systems, large deviations results have received a great deal of attention since their introduction in the 1980's [30, 37, 38, 12, 22, 44]. The results we study quantify the rate of convergence of empirical averages relative to a fixed reference measure  $m$ . That is, one studies the rate of decay of  $m\{x \mid \mathcal{E}_n(x) \in U\}$ , where  $\mathcal{E}_n(x)$  is the empirical measure of order  $n$  associated to the point  $x$ , and  $U$  is a suitable subset of the space of all probability measures on  $X$ . One may also consider level-1 large deviations principles, which study  $m\{x \mid \frac{1}{n}S_n\varphi(x) \in V\}$  for some fixed observable  $\varphi$  and some  $V \subset \mathbb{R}$ . Level-1 results for continuous observables follow from level-2 results via the contraction principle (see [10]).

To place our results in context, we summarize the main approaches to large deviations in dynamical systems. We recommend the introductions of [41, 27, 35, 10, 5] for further references and discussion. The Large Deviation Principle can be divided into two steps:

- (1) Upper Bound: bound  $\overline{\lim} \frac{1}{n} \log m\{x \mid \mathcal{E}_n(x) \in U\}$  from above in terms of a *rate function*. For an optimal result, we want to give the best possible description of the rate function.
- (2) Lower Bound: bound  $\underline{\lim} \frac{1}{n} \log m\{x \mid \mathcal{E}_n(x) \in U\}$  from below, ideally in terms of the same rate function as the upper bound.

It is often the case that the upper bound is easier to establish than the lower bound, requiring only an upper Gibbs bound on  $m$  (see §4.1). The task of obtaining a lower bound is generally carried out using one of the following three methods:

- (1) The ‘orbit-gluing approach’ relies on direct constructions based on the specification property (or one of its variants). This is the approach taken in this paper.
- (2) The ‘functional approach’ relies on differentiability of a certain functional on the space of observables, which can be related to uniqueness of equilibrium states.
- (3) The ‘tower approach’ relies on relating the original system to a countable state Markov shift via a tower construction. The

focus here is typically on the case where the reference measure is either volume or an SRB measure.

The first two approaches (orbit-gluing and functional) are fundamentally related to the theory of thermodynamic formalism, and have their full power in settings where this theory is well understood, such as uniformly hyperbolic systems. These approaches yield exponential rates of decay when they apply. The tower approach has been used to deal with a broad class of non-uniformly hyperbolic systems, where complete thermodynamic results are not always available and the rate of decay may be either exponential or polynomial.

We note that Araújo and Pacifico [1] have large deviations results based on the notion of *hyperbolic times*. These results apply to certain non-uniform and partially hyperbolic systems and do not fit into the classification above. We now recall some of the key results obtained in the literature using the three approaches.

*The tower approach.* The main advantages of the tower approach are its flexibility and the fact that it is the only method which yields results on both exponential and sub-exponential rates of decay. Quoting Rey-Bellet and Young [35], “The tower construction enables one to treat – in a unified way – a larger class of dynamical systems without insisting on optimal results.” The tower results of which we are aware focus on the case when the reference measure is either SRB or Lebesgue.

The tower approach dates back at least to Keller and Nowicki [21], who studied large deviations for Collet–Eckmann unimodal maps by using quasi-compactness of the transfer operator associated to a suitable tower. For Manneville–Pomeau maps, the level-1 large deviation principle for Hölder continuous observables was obtained by Pollicott, Sharp, and Yuri [34], in the cases where exponential decay rates apply, and level-2 results for both polynomial and exponential decay were later obtained by Pollicott and Sharp [33].

For a broader class of non-uniform examples, level-1 results for Hölder continuous observables have been obtained using the towers introduced by Young in [45]. Results for exponential decay were given by Rey-Bellet and Young in [35] and by Melbourne and Nicol in [27], using results on quasi-compact operators [19] to obtain local differentiability of the logarithmic moment generating function, whose Legendre transform gives the rate function. Using martingales, Melbourne and Nicol [27] also gave the first results for polynomial decay, which were improved by Melbourne in [26].

The results mentioned so far all have a ‘local’ aspect, which means that they describe the rate of decay of  $m\{x \mid \frac{1}{n}S_n\varphi(x) \in V\}$  only when

the set  $V$  is the complement of a sufficiently small neighbourhood of the expected value (the  $m$ -a.e. limit of  $\frac{1}{n}S_n\varphi$ ). ‘Global’ large deviations results, such as the ones we prove in this paper, do not hold without further conditions on the tower (see [4, §5] for examples). For towers satisfying an additional ‘nonsteep’ condition, Chung [4] has obtained a full level-2 large deviations principle, successfully removing both the ‘local’ assumption and the regularity assumption on the observables. He describes the rate function using Lyapunov exponents. These results have been applied to quadratic maps by Chung and Takahasi [5].

*The functional approach.* The technical and historical aspects of the functional approach are explained in detail in the introduction of [10]. This approach is similar in spirit to the original Gärtner–Ellis theorem [18, 14] from probability theory, and was first implemented in dynamical systems by Takahashi [37, 38]. A general formulation of the functional approach was given by Kifer [22]. The upper bound requires a weak version of the Gibbs property (such as [10, (1.5)]), and the key requirement for the lower bound is the existence of a dense subspace  $W \subset C(X)$  such that every  $\psi \in W$  has a unique equilibrium state.

The functional approach has been successfully applied to rational maps of the Riemann sphere. For hyperbolic rational maps, with the measure of maximal entropy as the reference measure, this was carried out by Lopes [25]. For the broader class of topological Collet–Eckmann rational maps, Comman and Rivera–Letelier [10] showed that every Hölder continuous potential  $\varphi$  has a unique equilibrium state  $\mu_\varphi$ , and the system satisfies level-2 large deviations with reference measure  $\mu_\varphi$ .

In the more abstract setting of this paper, so far there are no known axiomatic conditions to verify the dense subspace condition. This condition is known to hold for  $\beta$ -shifts [8, 42] and also follows for  $S$ -gap shifts from the arguments in §5 of this paper, but so far the proof in each case is specific to that particular example. It is a problem of independent interest to identify dynamical systems for which the dense subspace condition holds, although it may be challenging (or even impossible) to verify in some situations where we expect the large deviations results of this paper to apply.

*The orbit-gluing approach.* The idea of this method is that if we have the ability to ‘glue’ a collection of finite orbit segments into a single orbit segment, subject to a controlled amount of error, then we can obtain lower estimates via a constructive proof. The property which allows us to glue is called the specification property, and there are many variants on its precise definition, many of which are surveyed in [43]. The orbit-gluing approach to large deviations dates back at

least to work of Föllmer and Orey [17], who considered full  $\mathbb{Z}^d$ -shifts. The presence of phase transitions for  $d \geq 2$  renders the orbit-gluing approach preferable to the functional approach in this setting. Results for full  $\mathbb{Z}^d$ -shifts were also obtained by Eizenberg, Kifer, and Weiss [13].

The benchmark result in the orbit-gluing approach is that if a topological dynamical system  $(X, f)$  satisfies the specification property, and an invariant measure  $m$  on  $X$  satisfies the Gibbs property, then  $(X, f)$  satisfies the large deviations principle with reference measure  $m$ . A level-1 result in this form was given by Young [44, Theorem 1].

The specification property holds for uniformly hyperbolic dynamical systems, including topologically mixing Anosov diffeomorphisms and subshifts of finite type. In this setting, the Gibbs property can be verified for equilibrium measures corresponding to Hölder continuous potentials. The specification property and the Gibbs property are uniform global assumptions, and thus quite restrictive: in particular, they fail to hold for generic non-uniformly hyperbolic systems [28].

To apply the orbit-gluing method beyond this well understood setting, the challenge is to find appropriate conditions to replace the uniform ones. A key breakthrough in this direction is the work of Pfister and Sullivan [31], who used a weakened form of the specification property to prove that all  $\beta$ -shifts verify the large deviations principle (using the unique measure of maximal entropy as the reference measure). Using a similar approach, the third named author [43] used a suitable weak specification property to prove that ergodic automorphisms with positive topological entropy satisfy large deviations with Haar measure as the reference measure. The work of Varandas takes a similar philosophy (although using very different weak specification and Gibbs properties) in a smooth setting [41]. As far as we know, these are the only large deviations results which have been derived from weakened specification properties.

**Results in this paper.** The primary goal of this paper is to establish conditions under which a class of dynamical systems with non-uniform structure can be treated using the orbit-gluing approach. A significant difficulty in applying either the functional approach or the orbit-gluing approach in the non-uniform setting is to obtain the necessary thermodynamic results, in particular existence and uniqueness of equilibrium states and some version of the Gibbs property. The recent work of [7, 8] has provided the necessary thermodynamic groundwork to make this approach tractable for a large class of symbolic systems, by giving conditions for a potential  $\varphi$  to have a unique equilibrium state with the weak Gibbs property [A.3] – we state these in Theorem 3.1.

Our main result is that we can obtain a full level-2 large deviations principle from non-uniform versions of the Gibbs property and the specification property similar to those introduced in [7, 8].

We work in the symbolic setting because our motivating examples are symbolic and because the exposition and computations are simpler in this context. Our approach is suitable for adaptation to non-symbolic topological dynamical systems, where analogues of some of the ideas in this paper have already been introduced [9].

The criteria we introduce can be verified for many shift spaces in the class of *coded* systems, including  $S$ -gap shifts,  $\beta$ -shifts, and their factors. As shown in [7], there are  $S$ -gap shifts that do not satisfy any of the weak specification properties discussed in [43], including the almost specification property of [31, 40]. Thus, none of the previous large deviations results based on the ‘orbit-gluing’ approach can be applied to arbitrary  $S$ -gap shifts, so this is a completely new result for this family of shift spaces (we expect that existing tower methods would yield a ‘local’ level-1 result for arbitrary  $S$ -gap shifts, but this has never been carried out). For  $\beta$ -shifts, our result is a generalisation of [31], and the novel application here is that we are able to use a large class of reference measures, instead of just the measure of maximal entropy. For subshift factors of  $\beta$ -shifts and  $S$ -gap shifts, our result is completely new. Furthermore, the main results are formulated axiomatically; thus one may construct many more examples of shift spaces to which our results apply, and we expect that both the symbolic result and future generalisations to the non-symbolic setting will yield more applications in the future.

**Layout of the paper.** In §2, we establish our definitions. In §3, we give various results that follow from Theorem A by using the thermodynamic results developed in [7, 8]. In particular, we give applications to  $\beta$ -shifts,  $S$ -gap shifts, and their factors.

In §4, we prove Theorem A. The proofs of lemmas which are not proved in the body of the text appear in §6. In §5, we give proofs that the examples ( $\beta$ -shifts and  $S$ -gap shifts) satisfy the conditions of Theorem A. To apply Theorem A to every equilibrium state for a Bowen potential on an  $S$ -gap shift, we establish an intermediate result of independent interest which states that these equilibrium states always have positive entropy.

In an appendix, we fill in the details of the proof that (S)-specification can be replaced by (W)-specification in the main result of [8] (quoted here as Theorem 3.1), which is important for our examples.

## 2. DEFINITIONS

**2.1. Large deviations principles.** Let  $(X, d)$  be a compact metric space and  $f: X \rightarrow X$  a continuous map. Denote by  $\mathcal{M}(X)$  the set of all Borel probability measures on  $X$  with the weak\* topology. This topology is induced by the metric

$$D(\mu, \nu) := \sum_{n=1}^{\infty} \frac{|\int \varphi_n d\mu - \int \varphi_n d\nu|}{2^{n+1} \|\varphi_n\|_{\infty}},$$

where  $\{\varphi_n\} \subset C(X)$  is a countable dense subset. Let  $\mathcal{M}_f(X) \subset \mathcal{M}(X)$  be the set of  $f$ -invariant Borel probability measures, and let  $\mathcal{M}_f^e(X) \subset \mathcal{M}_f(X)$  be the set of ergodic measures. Given  $x \in X$ , consider the *empirical measures*

$$\mathcal{E}_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{f^j x},$$

where  $\delta_y$  is the point measure concentrated at  $y$ . For a fixed  $x$ , the sequence of empirical measures  $\mathcal{E}_n(x)$  converges to  $\mathcal{M}_f(X)$ . Large deviations theory quantifies some aspects of that convergence.

**Definition 2.1.** We say that the system  $(X, f)$  satisfies a *level-2 large deviations principle* with a reference measure  $m \in \mathcal{M}(X)$  and a rate function  $q: \mathcal{M}(X) \rightarrow [-\infty, 0]$  if  $q$  is upper semicontinuous,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X \mid \mathcal{E}_n(x) \in U\}) \geq \sup_{\mu \in U} q(\mu)$$

holds for any open set  $U \subset \mathcal{M}(X)$ , and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X \mid \mathcal{E}_n(x) \in F\}) \leq \sup_{\mu \in F} q(\mu)$$

holds for any closed set  $F \subset \mathcal{M}(X)$ .

**2.2. Shift spaces and languages.** Let  $A$  be a finite set and  $A^{\mathbb{N}}$  (resp.  $A^{\mathbb{Z}}$ ) be the set of all one-sided (resp. two-sided) infinite sequences on the alphabet  $A$ , endowed with the standard metric  $d(x, y) = 2^{-t(x, y)}$ , where  $t(x, y) = \min\{|k| \mid x_k \neq y_k\}$ . The shift map on  $A^{\mathbb{N}}$  is  $\sigma: x_1 x_2 \cdots \mapsto x_2 x_3 \cdots$ , and the shift map on  $A^{\mathbb{Z}}$  is defined analogously. A subshift is a closed  $\sigma$ -invariant set  $X \subset A^{\mathbb{N}}$  or  $X \subset A^{\mathbb{Z}}$ . All of the results and proofs in this paper apply equally to one-sided and two-sided shifts, so we treat both cases simultaneously.

The *language of  $X$* , denoted by  $\mathcal{L} = \mathcal{L}(X)$ , is the set of all finite words that appear in any sequence  $x \in X$  – that is,

$$\mathcal{L}(X) = \{w \in A^* \mid [w] \neq \emptyset\},$$



where  $A^* = \bigcup_{n \geq 0} A^n$  and  $[w]$  is the central cylinder for  $w$ , which in the one-sided case is the set of sequences  $x \in X$  that begin with the word  $w$ . Given  $w \in \mathcal{L}$ , let  $|w|$  denote the length of  $w$ . For any collection  $\mathcal{D} \subset \mathcal{L}$ , let  $\mathcal{D}_n$  denote  $\{w \in \mathcal{D} \mid |w| = n\}$ . Thus,  $\mathcal{L}_n$  is the set of all words of length  $n$  that appear in sequences belonging to  $X$ . Given words  $u, v$ , we use juxtaposition  $uv$  to denote the word obtained by concatenation.

A *decomposition* of  $\mathcal{L}$  is a collection of three sets of words  $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$  such that given any  $w \in \mathcal{L}$ , there exist  $u^p \in \mathcal{C}^p, v \in \mathcal{G}, u^s \in \mathcal{C}^s$  such that  $w = u^p v u^s$ . We write  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  when the language can be decomposed in this way. We make a standing assumption that  $\emptyset \in \mathcal{C}^p, \mathcal{G}, \mathcal{C}^s$  to allow for words in  $\mathcal{L}$  that belong purely to one of the three collections (this is also implicit in [7, 8]).

Once a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  has been fixed, we consider for each  $M \in \mathbb{N}$  the set

$$(2.1) \quad \mathcal{G}^M := \{uvw \in \mathcal{L} \mid v \in \mathcal{G}, u \in \mathcal{C}^p, w \in \mathcal{C}^s, |u| \leq M, |w| \leq M\}.$$

Note that  $\mathcal{L} = \bigcup_{M \in \mathbb{N}} \mathcal{G}^M$ , so this defines a filtration of the language of  $X$ .

**2.3. Entropy and pressure for shift spaces.** Given a collection  $\mathcal{D} \subset \mathcal{L}$ , the *entropy* of  $\mathcal{D}$  is

$$h(\mathcal{D}) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \# \mathcal{D}_n,$$

where  $\mathcal{D}_n = \{w \in \mathcal{D} \mid |w| = n\}$ . The entropy of an invariant measure  $\mu \in \mathcal{M}_\sigma(X)$  is

$$h(\mu) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{w \in \mathcal{L}_n} -\mu[w] \log \mu[w].$$

For a fixed potential function  $\varphi \in C(X)$ , the *pressure* of  $\mathcal{D} \subset \mathcal{L}$  is

$$P(\mathcal{D}, \varphi) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \Lambda_n(\mathcal{D}, \varphi),$$

where

$$\Lambda_n(\mathcal{D}, \varphi) = \sum_{w \in \mathcal{D}_n} e^{\sup_{x \in [w]} S_n \varphi(x)}$$

and  $S_n \varphi(x) = \sum_{k=0}^{n-1} \varphi(\sigma^k x)$ . We write  $P(\varphi) = P(\mathcal{L}, \varphi)$ .

We will be primarily concerned with potentials having some extra regularity: we say that  $\varphi$  has the *Bowen property on  $\mathcal{D}$*  if there is  $V \in \mathbb{R}$  such that for every  $n \in \mathbb{N}$ , every  $w \in \mathcal{D}_n$ , and every  $x, y \in [w]$ , we have  $|S_n \varphi(x) - S_n \varphi(y)| \leq V$ . In particular, if  $\varphi$  is Hölder continuous then it has the Bowen property on every  $\mathcal{D} \subset \mathcal{L}$ .



**2.4. Specification.** We define the specification properties that appear in this paper, and the relationships between them. In [7], we introduced the following definitions.

**Definition 2.2.** Given a shift space  $X$  and its language  $\mathcal{L}$ , consider a subset  $\mathcal{G} \subset \mathcal{L}$ . Fix  $\tau \in \mathbb{N}$ ; any of the following conditions defines a *specification property on  $\mathcal{G}$*  with gap size  $\tau$ .

- (W): For all  $m \in \mathbb{N}$  and  $w^1, \dots, w^m \in \mathcal{G}$ , there exist  $v^1, \dots, v^{m-1} \in \mathcal{L}$  such that  $x := w^1 v^1 w^2 v^2 \dots v^{m-1} w^m \in \mathcal{L}$  and  $|v^i| \leq \tau$  for all  $i$ .
- (S): Condition (W) holds, and in addition, the connecting words  $v^i$  can all be chosen to have length exactly  $\tau$ .

In the case  $\mathcal{G} = \mathcal{L}$ , these definitions are equivalent to the well known weak specification, respectively specification, property of the shift.

In this paper, we also consider the following specification properties.

**Definition 2.3.** Given a shift space  $X$  and its language  $\mathcal{L}$ , consider a subset  $\mathcal{G} \subset \mathcal{L}$ , we say that  $\mathcal{G}$  has *(W)-specification with good concatenations*, or simply *(gcW)-specification* if for all  $m \in \mathbb{N}$  and  $w^1, \dots, w^m \in \mathcal{G}$ , there exist  $v^1, \dots, v^{m-1} \in \mathcal{L}$  such that

- (1)  $|v^i| \leq \tau$  for all  $i$ , and
- (2)  $w^i v^i \dots v^{j-1} w^j \in \mathcal{G}$  for every  $1 \leq i < j \leq m$ .

The meaning of (gcW)-specification is essentially that words in  $\mathcal{G}$  can always be glued together to give new words in  $\mathcal{G}$ .

**Definition 2.4.** Given a shift space  $X$  and its language  $\mathcal{L}$ , consider a subset  $\mathcal{G} \subset \mathcal{L}$ . We say that  $\mathcal{G}$  has *specification with transition time 0*, or (0)-specification, if for all  $u, w \in \mathcal{G}$ , we have  $uw \in \mathcal{G}$ . We may also refer to (0)-specification as the *free concatenation* property.

It is clear from the definitions that (0)-specification implies (gcW)-specification, and that (gcW)-specification implies (W)-specification. The advantage of (0)-specification is that it is a simple criterion, which can be verified easily for natural collections  $\mathcal{G}$  associated with  $\beta$ -shifts,  $S$ -gap shifts, and more general coded systems.

**2.5. Edit Approachability.** First we introduce the edit metric (sometimes known as the Damerau–Levenshtein metric) on  $\mathcal{L}$ .

**Definition 2.5.** Define an *edit* of a word  $w = w_1 \dots w_n \in \mathcal{L}$  to be a transformation of  $w$  by one of the following actions, where  $u^j \in \mathcal{L}$  are arbitrary words and  $a, a' \in A$  are arbitrary symbols.

- (1) *Substitution*:  $w = u^1 a u^2 \mapsto w' = u^1 a' u^2$ .
- (2) *Insertion*:  $w = u^1 u^2 \mapsto w' = u^1 a' u^2$ .

(3) *Deletion*:  $w = u^1 a u^2 \mapsto w' = u^1 u^2$ .

Given  $v, w \in \mathcal{L}$ , define the *edit distance* between  $v$  and  $w$  to be the minimum number of edits required to transform the word  $v$  into the word  $w$ : we will denote this by  $\hat{d}(v, w)$ .

The following lemma about the size of balls in the edit metric will be crucial for our entropy estimates.

**Lemma 2.6.** *There is  $C > 0$  such that given  $n \in \mathbb{N}$ ,  $w \in \mathcal{L}_n$ , and  $\delta > 0$ , we have*

$$(2.2) \quad \#\{v \in \mathcal{L} \mid \hat{d}(v, w) \leq \delta n\} \leq C n^C (e^{C\delta} e^{-\delta \log \delta})^n.$$

Now we can introduce our key new definition, which requires that any word in  $\mathcal{L}$  can be transformed into a word in  $\mathcal{G}$  with a relatively small number of edits.

**Definition 2.7.** Say that a non-decreasing function  $g: \mathbb{N} \rightarrow \mathbb{N}$  is a *mistake function* if  $\frac{g(n)}{n}$  converges to 0. We say that  $\mathcal{L}$  is *edit approachable* by  $\mathcal{G}$ , where  $\mathcal{G} \subset \mathcal{L}$ , if there is a mistake function  $g$  such that for every  $w \in \mathcal{L}$ , there exists  $v \in \mathcal{G}$  with  $\hat{d}(v, w) \leq g(|w|)$ .

An important consequence of edit approachability is that we can replace sufficiently long words in  $\mathcal{L}$  with words in  $\mathcal{G}$  in such a way that estimates on Birkhoff averages, and thus estimates on empirical measures, can be well controlled, while at the same time, (2.2) guarantees that not much entropy is lost this way.

Control on the Birkhoff averages is given by the following lemma.

**Lemma 2.8.** *For any continuous function  $\varphi: X \rightarrow \mathbb{R}$  and any mistake function  $g(n)$ , there is a sequence of positive numbers  $\delta_n \rightarrow 0$  such that if  $x, y \in X$  and  $m, n \in \mathbb{N}$  are such that  $\hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq g(n)$ , then  $|\frac{1}{n} S_n \varphi(x) - \frac{1}{m} S_m \varphi(y)| \leq \delta_n$ .*

It follows from Lemmas 2.6 and 2.8 that edit approachability implies that the collection  $\mathcal{G}$  carries full pressure for every continuous potential (this result is included for independent interest although we do not use it in this paper).

**Proposition 2.9.** *If  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , then  $P(\mathcal{G}, \varphi) = P(\varphi)$  for every  $\varphi \in C(X)$ .*

**Remark 2.10.** If, in Definition 2.7, we replace  $\hat{d}(v, w)$  with the Hamming distance  $d_{\text{Ham}}$  between  $v$  and  $w$ , we could say that  $\mathcal{L}$  is *Hamming approachable* by  $\mathcal{G}$ . Clearly, if  $\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ , then  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ . If  $\mathcal{L}$  is Hamming approachable

by  $\mathcal{G}$ , and  $\mathcal{G}$  satisfies the specification property, then it is easy to see that the symbolic space satisfies the *almost specification* property of [7, §3.3]: there exists a mistake function  $g$  so that for any words  $u, v \in \mathcal{L}$ , there are words  $u', v' \in \mathcal{L}$  so that  $u'v' \in \mathcal{L}$ ,  $d_{\text{Ham}}(u, u') \leq g(|u|)$  and  $d_{\text{Ham}}(v, v') \leq g(|v|)$ .

Hamming approachability is too strong an assumption for our application to  $S$ -gap shifts, since there are examples of  $S$ -gap shifts without the almost specification property [7, §3.3]. When  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , and  $\mathcal{G}$  has (W)-specification, we see that the symbolic space satisfies a weaker version of almost specification, given by replacing Hamming distance with edit distance in the definition of almost specification.

**2.6. Gibbs properties.** The standard Gibbs property for a shift space says that a measure  $m \in \mathcal{M}(X)$  is *Gibbs* if there are constants  $K, K' > 0$  such that

$$(2.3) \quad K \leq \frac{m[x_1 \cdots x_n]}{e^{-nP(\varphi) + S_n \varphi(x)}} \leq K'$$

for all  $x \in X$ . We will require that the upper bound hold uniformly, while the lower bound will only be required to hold when  $x_1 \cdots x_n \in \mathcal{G}$ . More precisely, we make the following definition for a collection  $\mathcal{G} \subset \mathcal{L}$ .

**Definition 2.11.** A measure  $m \in \mathcal{M}(X)$  is *Gibbs with respect to  $\mathcal{G}$*  if there are constants  $K, K' > 0$  such that

$$(2.4) \quad m[x_1 \cdots x_n] \leq K' e^{-nP(\varphi) + S_n \varphi(x)}$$

for every  $x \in X$  and  $n \in \mathbb{N}$ , and

$$(2.5) \quad m[x_1 \cdots x_n] \geq K e^{-nP(\varphi) + S_n \varphi(x)}$$

whenever  $x \in X$  and  $n \in \mathbb{N}$  are such that  $x_1 \cdots x_n \in \mathcal{G}$ .

Definition 2.11 is property [A.3] of Theorem A. Theorem 3.1 provides examples of measures satisfying this definition.

**2.7. Properties under factors.** One advantage of our techniques is that they behave well under factors. We let  $\Sigma$  be a shift space, and we let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$ . Suppose that  $X$  is a topological factor of  $\Sigma$ , that is, there exists a continuous surjective map  $\pi: \Sigma \rightarrow X$  such that  $\sigma \circ \pi = \pi \circ \sigma$ . By [24, Theorem 6.29],  $\pi$  is a block code: there exist  $r \in \mathbb{N}$  and  $\psi: \mathcal{L}_{2r+1} \rightarrow \mathcal{A}$ , where  $\mathcal{A}$  is the alphabet of  $X$ , such that

$$(2.6) \quad (\pi x)_n = \psi(x_{n-r} x_{n-r+1} \cdots x_{n+r-1} x_{n+r}).$$

This induces a surjective map  $\Psi: \mathcal{L}(\Sigma)_{n+2r} \rightarrow \mathcal{L}(X)_n$  by

$$\Psi(w_1 \cdots w_{n+2r}) = \psi(w_1 \cdots w_{2r+1}) \psi(w_{2r+2} \cdots w_{2r+2r}) \cdots \psi(w_n \cdots w_{n+2r}).$$

We set  $\tilde{\mathcal{G}} = \Psi(\mathcal{G})$ . The key to our study of  $X$  is that  $\tilde{\mathcal{G}}$  inherits a number of good properties of  $\mathcal{G}$ , including in particular [A.1], [A.2], and the condition (I) that appears in Theorem 3.1 below.

**Lemma 2.12.** *Let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$  and  $\tilde{\mathcal{G}} \subset \mathcal{L}(X)$  be as above.*

- (1) *If  $\mathcal{G}$  satisfies [A.1], then  $\tilde{\mathcal{G}}$  satisfies [A.1].*
- (2) *If  $\mathcal{G}$  satisfies [A.2], then  $\tilde{\mathcal{G}}$  satisfies [A.2].*
- (3) *If  $\mathcal{G}$  satisfies (I), then  $\tilde{\mathcal{G}}$  satisfies (I).*

Furthermore, if  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s$  is a decomposition for  $\mathcal{L}(\Sigma)$ , there is a natural decomposition for  $\mathcal{L}(X)$ . We define  $\tilde{\mathcal{C}}^p$  by taking  $\Psi(\mathcal{C}^p \mathcal{L}_{2k}(\Sigma))$ , and  $\tilde{\mathcal{C}}^s$  by taking  $\Psi(\mathcal{L}_{2k}(\Sigma) \mathcal{C}^s)$ . It is easy to check the following lemma.

**Lemma 2.13.** *If  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s$  is a decomposition for  $\mathcal{L}(\Sigma)$ , then  $\tilde{\mathcal{C}}^p \tilde{\mathcal{G}} \tilde{\mathcal{C}}^s$  is a decomposition for  $\mathcal{L}(X)$ . If  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ , then  $h(\tilde{\mathcal{C}}^p \cup \tilde{\mathcal{C}}^s) = 0$ .*

### 3. CONSEQUENCES OF THEOREM A

**3.1. Unique equilibrium states.** The following result from [7, 8] provides unique equilibrium states which satisfy the weak Gibbs property [A.3], and is our key tool for finding reference measures to which Theorem A applies. Roughly speaking, Conditions (I)–(II) state that  $\mathcal{C}^p$  and  $\mathcal{C}^s$  contain all obstructions to specification (for the system) and regularity (for the potential), while Condition (III) states that these obstructions carry smaller pressure than the whole system.

**Theorem 3.1** ([8], Theorem C and Remark 2.2). *Let  $(X, \sigma)$  be a subshift on a finite alphabet and  $\varphi \in C(X)$  a potential. Suppose there exist collections of words  $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$  such that  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s = \mathcal{L}$  and the following conditions hold:*

- (I)  $\mathcal{G}^M$  has  $(W)$ -specification for every  $M \in \mathbb{N}$ ;
- (II)  $\varphi$  has the Bowen property on  $\mathcal{G}$ ;
- (III)  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$ .

*Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and  $m_\varphi$  is Gibbs with respect to  $\mathcal{G}$ . In particular,  $m_\varphi$  satisfies [A.3].*

In [8], the stronger condition of (S)-specification is assumed in (I), but the proof goes through with only minor modifications, which we present these in Appendix A. Combining Theorem A and Theorem 3.1, we immediately obtain the following result.

**Theorem B.** *Let  $X$  be a subshift on a finite alphabet and  $\varphi \in C(X)$  a potential. Suppose  $\mathcal{L}$  has a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  satisfying [A.1], [A.2] and (I)–(III). Then writing  $m_\varphi$  for the unique equilibrium state*

of  $\varphi$ , the system  $(X, \sigma)$  satisfies a level-2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q^\varphi$  given by (1.1).

For a shift space  $X$  and a collection of words  $\mathcal{C} \subset \mathcal{L}$ , it is typically much easier to verify  $h(\mathcal{C}) < h(X)$  than  $P(\mathcal{C}, \varphi) < P(X, \varphi)$ . For  $\beta$ -shifts, it was shown in [8, Proposition 3.1] that **(III)** holds for every Bowen potential  $\varphi$ , and we show in §5 that this is also true for  $S$ -gap shifts. However, for other shift spaces where no analogous argument is available yet, the following lemma is a convenient way to ensure that **(III)** holds for a large class of functions.

**Lemma 3.2.** *Suppose  $X$  is a shift space and  $\mathcal{C} \subset \mathcal{L}(X)$  is a collection of words such that  $h(\mathcal{C}) < h(X)$ , and let  $\varphi: X \rightarrow \mathbb{R}$ . If  $\varphi$  satisfies the bounded range condition*

$$(BR) \quad \sup \varphi - \inf \varphi < h(X) - h(\mathcal{C}),$$

*then  $P(\mathcal{C}, \varphi) < P(X, \varphi)$ .*

Often we can take  $h(\mathcal{C}) = 0$ , in which case the condition (BR) on  $\varphi$  reduces to the condition

$$(BR_0) \quad \sup \varphi - \inf \varphi < h(X).$$

**3.2. Factors.** Our results are well behaved under the operation of passing to a subshift factor and we have the following general theorem.

**Theorem C.** *Let  $\Sigma$  be a subshift on a finite alphabet, and suppose that  $\mathcal{L} = \mathcal{L}(\Sigma)$  has a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  satisfying **[A.1]**, **[A.2]** and **(I)**. Assume further that  $h(\mathcal{C}) = 0$ , where  $\mathcal{C} = \mathcal{C}^p \cup \mathcal{C}^s$ . Let  $X$  be a subshift factor of  $\Sigma$  and  $\varphi: X \rightarrow \mathbb{R}$  be a continuous function satisfying  $(BR_0)$  and the Bowen property. Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and the system  $(X, \sigma)$  satisfies a level-2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q^\varphi$  given by (1.1).*

**3.3. Coded systems.** Our examples all fall into the class of coded systems, which are shift spaces whose language can be obtained by freely concatenating a countable collection of generating words  $G$ . That is,  $X$  is a coded system if there exists a countable collection of finite words  $G$  such that the language  $\mathcal{L}(X)$  is characterised as follows: a word  $v$  is contained in  $\mathcal{L}(X)$  if and only if there are  $w^1, \dots, w^n \in G$  such that  $v$  is a subword of  $w^1 \dots w^n$ .

For coded shifts we may take  $\mathcal{G}$  to be the set of finite concatenations of generators from  $G$ . Then **[A.1]** is automatically satisfied and the brief argument of [7, §4] shows that **(I)** holds.

**Theorem D.** *Let  $X$  be a coded system,  $m$  a Borel probability measure on  $X$ , and  $\varphi: X \rightarrow \mathbb{R}$  a continuous function. Let  $G \subset \mathcal{L}(X)$  be a collection of generators for  $X$  and let  $\mathcal{G}$  be the set of finite concatenations of words in  $G$ . If [A.2] and [A.3] are satisfied, then  $(X, \sigma)$  satisfies a level-2 large deviation principle with reference measure  $m$  and rate function  $q^\varphi$  given by (1.1).*

The following result follows immediately from Theorem 3.1, Theorem D, and Lemma 3.2.

**Theorem E.** *Let  $(X, \sigma)$  be a coded system with generating set  $G$ . Let  $\mathcal{G}$  be the collection of finite concatenations of generators, and let  $\mathcal{C}$  be the collection of prefixes and suffixes of generators. If  $h(\mathcal{C}) < h(X)$  and if  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , then every potential  $\varphi$  with the Bowen property and the bounded range condition (BR) has a unique equilibrium state satisfying the weak Gibbs property [A.3] and the large deviations principle in Theorem A.*

**3.4.  $\beta$ -shifts and  $S$ -gap shifts.** Our main examples are the  $\beta$ -shifts, the  $S$ -gap shifts, and their subshift factors. For all of these examples we can take  $h(\mathcal{C}) = 0$ , and for the  $\beta$ -shifts (in [8, Proposition 3.1]) and  $S$ -gap shifts (in §5.1), we can show that  $P(\mathcal{C}, \varphi) < P(\varphi)$  for all Bowen potentials  $\varphi$ , which removes the need for the bounded range condition. For factors of  $\beta$ -shifts and  $S$ -gap shifts, we do require the additional assumption (BR<sub>0</sub>) on the potential  $\varphi$  at present.

**3.4.1.  $S$ -gap shifts.** An  $S$ -gap shift  $\Sigma_S$  is a subshift of  $\{0, 1\}^{\mathbb{Z}}$  defined by the rule that for a fixed  $S \subset \{0, 1, 2, \dots\}$ , the number of 0's between consecutive 1's is an integer in  $S$ . More precisely, the language of  $\Sigma_S$  is

$$\{0^n 10^{n_1} 10^{n_2} 1 \dots 10^{n_k} 10^m \mid n_i \in S \text{ for all } 1 \leq i \leq k \text{ and } n, m \in \mathbb{N}\},$$

together with  $\{0^n \mid n \in \mathbb{N}\}$ , where we assume that  $S$  is infinite (when  $S$  is finite,  $\Sigma_S$  is sofic and can be analysed without the techniques of this paper). The language for  $\Sigma_S$  admits the following decomposition:

$$\begin{aligned} \mathcal{G} &= \{0^{n_1} 10^{n_2} 10^{n_3} 1 \dots 10^{n_k} 1 \mid n_i \in S \text{ for all } 1 \leq i \leq k\}, \\ \mathcal{C}^p &= \{0^n 1 \mid n \notin S\}, \\ \mathcal{C}^s &= \{0^n \mid n \in \mathbb{N}\}, \end{aligned}$$

which was first studied in [7]. We verify in §5.1 that this decomposition satisfies Conditions [A.1], [A.2], and (I)–(III) for every Bowen potential  $\varphi$ .

3.4.2.  *$\beta$ -shifts.* Fix  $\beta > 1$ , write  $b = \lceil \beta \rceil$ , and let  $\omega^\beta \in \{0, 1, \dots, b-1\}^\mathbb{N}$  be the greedy  $\beta$ -expansion of 1. Then  $\omega^\beta$  satisfies

$$\sum_{j=1}^{\infty} \omega_j^\beta \beta^{-j} = 1,$$

and has the property that  $\sigma^j(\omega^\beta) \preceq \omega^\beta$  for all  $j \geq 1$ , where  $\preceq$  denotes the lexicographic ordering. The  $\beta$ -shift is defined by

$$\Sigma_\beta = \{x \in \{0, 1, \dots, b-1\}^\mathbb{N} \mid \sigma^j(x) \preceq \omega^\beta \text{ for all } j \geq 1\}.$$

The first and second author showed in [7, 8] that the language for  $\Sigma_\beta$  admits a decomposition  $\mathcal{L}(\Sigma_\beta) = \mathcal{GC}^s$  that satisfies **(I)**–**(III)** for every Bowen potential  $\varphi$ . In §5.2 we briefly review the construction and show that conditions **[A.1]** and **[A.2]** are also satisfied.

3.4.3. *Results for examples.* We collect our results as applied to these examples in the following theorem. We say that a subshift is non-trivial if it is not a single periodic orbit. We proved in [7, Proposition 2.4] that a non-trivial subshift factor of a  $\beta$ -shift or  $S$ -gap shift has positive entropy.

**Theorem F.** *Let  $X$  and  $\varphi$  be one of the following:*

- (1)  *$X$  is a  $\beta$ -shift or an  $S$ -gap shift, and  $\varphi$  has the Bowen property;*
- (2)  *$X$  is a non-trivial subshift factor of a  $\beta$ -shift or an  $S$ -gap shift, and  $\varphi$  satisfies  $(\text{BR}_0)$  and the Bowen property.*

*Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and  $(X, \sigma)$  satisfies a level 2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q: \mathcal{M}(X) \rightarrow [-\infty, 0]$  given by*

$$q(\mu) = \begin{cases} h(\mu) + \int \varphi d\mu - P(\varphi) & \mu \text{ is } \sigma\text{-invariant,} \\ -\infty & \text{otherwise.} \end{cases}$$

*In particular, taking for our reference measure the unique measure of maximal entropy  $m_0$  (since  $\varphi \equiv 0$  always satisfies  $(\text{BR}_0)$ ), the system satisfies a level 2 large deviation principle with rate function given by*

$$q(\mu) = \begin{cases} h(\mu) - h_{\text{top}}(X, \sigma) & \mu \text{ is } \sigma\text{-invariant,} \\ -\infty & \text{otherwise.} \end{cases}$$

To the best of our knowledge, the only statement above which was previously known is the case when  $X$  is a  $\beta$ -shift and  $m_0$  is the measure of maximal entropy [31] (apart from the exceptional set of special cases above where  $X$  has specification).



**3.5. A ‘horseshoe’ theorem.** We now state a result that we establish as a key step in the proof of Theorem A, which may be of independent interest.

**Proposition 3.3.** *Let  $X$  be a shift space and suppose that  $\mathcal{G} \subset \mathcal{L}$  satisfies [A.1] and [A.2]. Then there exists an increasing sequence  $\{X_n\}$  of compact  $\sigma$ -invariant subsets of  $X$  with the following properties.*

- (1) *For every  $n$  and every  $w \in \mathcal{L}(X_n)$ , there exist  $u, v \in \mathcal{L}$  with  $|u|, |v| \leq n + \tau$  such that  $uwv \in \mathcal{G}$ . In particular, this implies that each  $X_n$  has the  $(W)$ -specification property.*
- (2) *Every invariant measure on  $X$  is entropy approachable by ergodic measures on  $X_n$ : for any  $\eta > 0$ , any  $\mu \in \mathcal{M}_\sigma(X)$ , and any neighborhood  $U$  of  $\mu$  in  $\mathcal{M}_\sigma(X)$ , there exist  $n \geq 1$  and  $\mu' \in \mathcal{M}_\sigma^e(X_n) \cap U$  such that  $h(\mu') > h(\mu) - \eta$  holds.*

By the variational principle and the entropy approachability in Proposition 3.3, we have the further result that  $\lim_{n \rightarrow \infty} h(X_n) = h(X)$ , and more generally

$$P(X, \varphi) = \lim_{n \rightarrow \infty} P(X_n, \varphi) = \sup_{n \in \mathbb{N}} P(X_n, \varphi)$$

for every  $\varphi \in C(X)$ . Thus, we can interpret the sets  $X_n$  as well behaved ‘horseshoes’ which can be used to approximate the original space  $X$ , revealing a structure reminiscent of Katok horseshoes [20]. Similarly, the filtration  $\mathcal{L} = \bigcup_{M \in \mathbb{N}} \mathcal{G}^M$  of the language of the shift can be considered to be analogous to the filtration of a non-uniformly hyperbolic set into Pesin sets.

**Remark 3.4.** In the case of coded systems, the subshifts  $X_n$  also satisfy  $\overline{\bigcup_n X_n} = X$  (see Lemma 4.5).

In the proof of the main results, we will use the following consequence of the first property in Proposition 3.3: If a measure  $m \in \mathcal{M}(X)$  is Gibbs with respect to  $\mathcal{G}$ , then  $m$  has the following Gibbs property on the family of subshifts  $\{X_n\}$ : there exist constants  $K_n, K' > 0$  such that for every  $x \in X_n$  and  $k \in \mathbb{N}$ , we have

$$(3.1) \quad K_n \leq \frac{m[x_1 \cdots x_k]}{e^{-kP(\varphi) + S_k \varphi(x)}} \leq K'.$$

This follows from the fact that  $x_1 \cdots x_k$  can be extended to a word in  $\mathcal{G}$  by adding a word to each end whose length is bounded by a constant depending only on  $n$ .

## 4. PROOF OF THEOREM A

The large deviations property in Definition 2.1 comprises an upper bound and a lower bound. We establish the upper bound first, then prove Proposition 3.3, which is the key to establishing the lower bound. For the upper bound, we use criteria given by Pfister and Sullivan in [31].

**4.1. Upper bound.** Given  $\mu \in \mathcal{M}_\sigma(X)$ , let  $q^\varphi(\mu) = h(\mu) + \int \varphi d\mu - P(\varphi)$ , as in (1.1). We show that for any closed set  $F \subset \mathcal{M}(X)$ , we have

$$(4.1) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log m(\mathcal{E}_n^{-1}(F)) \leq \sup_{\mu \in F \cap \mathcal{M}_\sigma(X)} q^\varphi(\mu).$$

Our key tool is the following result, which is a combination of Theorem 3.2 and Proposition 4.2 of [31].

**Theorem 4.1.** [31, Theorem 3.2 and Proposition 4.2] *Let  $(X, \sigma)$  be a subshift,  $m \in \mathcal{M}(X)$ ,  $\psi \in C(X)$ , and assume that the equation*

$$(4.2) \quad \limsup_{n \rightarrow \infty} \sup_{w \in \mathcal{L}_n} \left( \frac{1}{n} \log m([w]) + \frac{1}{n} \sup_{x \in [w]} S_n \psi(w) \right) \leq 0$$

*holds. Then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log m(\mathcal{E}_n^{-1}(F)) \leq \sup_{\mu \in F \cap \mathcal{M}_\sigma(X)} \left( h(\mu) - \int \psi d\mu \right).$$

We will apply Theorem 4.1 with  $\psi = P(\varphi) - \varphi$ . The upper Gibbs bound in [A.3] (see (2.4)) yields a constant  $K'$  such that

$$m([w]) \leq K' e^{-nP(\varphi) + S_n \varphi(x)}$$

for every  $x \in [w]$ . Thus

$$\frac{1}{n} \log m([w]) + \frac{1}{n} \sup_{x \in [w]} S_n (P(\varphi) - \varphi)(x) \leq \frac{1}{n} \log(K') \rightarrow 0,$$

for every  $x \in [w]$ . This establishes (4.2), so Theorem 4.1 provides the desired upper bound.

**4.2. The gluing map.** The specification property allows us to define the following gluing map, which we consider in both Lemma 4.10 and Appendix A. Given a collection of words  $\mathcal{D} \subset \mathcal{L}$  with (W)-specification, we write  $\mathcal{D}^*$  for the set of all finite sequences  $(w^1, \dots, w^m)$  where each  $w^i \in \mathcal{D}$ . For each  $(w^1, \dots, w^m) \in \mathcal{D}^*$ , we use (W)-specification to choose a word

$$(4.3) \quad \Phi(w^1, \dots, w^m) := w^1 v^1 \dots v^{m-1} w^m,$$

so that  $v^1, \dots, v^{m-1}$  are words with  $|v^i| \leq \tau$  and  $\Phi(w^1, \dots, w^m) \in \mathcal{L}$ . Selecting such a word for each  $(w^1, \dots, w^m) \in \mathcal{D}^*$  defines a map  $\Phi: \mathcal{D}^* \rightarrow \mathcal{L}$  which we call the *gluing map*. The map  $\Phi$  extends from  $\mathcal{D}^*$  to  $\mathcal{D}^{\mathbb{N}}$  in the natural way. Note that if  $\mathcal{D}$  satisfies the stronger assumption of the (gcW)-specification property, then each of the words  $\Phi(w^1, \dots, w^m)$  also belong to  $\mathcal{D}$ .

When  $\mathcal{D}$  has (S)-specification, we are able to further require that the words  $v^1, \dots, v^{m-1}$  in (4.3) have equal length, and it is obvious that for any choice of  $n_1, \dots, n_k$ , the restriction of the gluing map  $\Phi$  to  $\prod_{i=1}^k \mathcal{D}_{n_i}$  is injective.

When  $\mathcal{D}$  only has (W)-specification, we need to handle the possibility that  $\Phi$  may fail to be injective on  $\prod_{i=1}^k \mathcal{D}_{n_i}$ . Because of the possibility that the gluing words  $v^i$  may vary in length, we observe that the words  $\Phi(\vec{w})$  may have different lengths for different choices of  $\vec{w}$ , and so we work with the following *truncated* gluing map  $\Phi_0: \mathcal{D}^* \rightarrow \mathcal{L}$ .

**Definition 4.2.** For each  $n_1, \dots, n_k \in \mathbb{N}$ , the *truncated gluing map*  $\Phi_0$  on  $\prod_{i=1}^k \mathcal{D}_{n_i}$  is the map which takes  $(w^1, \dots, w^m) \in \prod_{i=1}^k \mathcal{D}_{n_i}$  to  $\Phi(w^1, \dots, w^m)$ , and then truncates to the first  $\sum_{i=1}^k n_i$  symbols. That is, writing  $N = \sum_{i=1}^k n_i$ , the map  $\Phi_0: \prod_{i=1}^k \mathcal{D}_{n_i} \rightarrow \mathcal{L}_N$  is given by  $i_N^p \circ \Phi$ , where  $i_N^p$  is the map on words of length at least  $N$  which truncates to the first  $N$  symbols.

**Lemma 4.3.** Suppose  $\mathcal{D} \subset \mathcal{L}$  has (W)-specification. There exists  $C > 0$  such that for any  $n_1, \dots, n_k \in \mathbb{N}$ , the truncated gluing map  $\Phi_0: \prod_{i=1}^k \mathcal{D}_{n_i} \rightarrow \mathcal{L}_N$  satisfies  $\#\Phi_0^{-1}(v) \leq C^k$  for each  $v \in \mathcal{L}_N$ , where  $N = \sum_{i=1}^k n_i$ .

*Proof.* Define  $\vec{\ell}: \mathcal{D}^* \rightarrow \{0, 1, \dots, \tau\}^*$  by  $\vec{\ell}(\vec{w}) = (|v^1|, |v^2|, \dots, |v^{k-1}|)$ , where  $v_i$  are the gluing words from the definition of  $\Phi$ . We consider  $\Phi_0: \prod_{i=1}^k \mathcal{D}_{n_i} \rightarrow \mathcal{L}$ . The image of  $\Phi_0$  is a subset of  $\mathcal{L}_N$ , where  $N = \sum_{i=1}^k n_i$ , so we choose  $v \in \mathcal{L}_N$ .

We fix  $\vec{\tau} = (\tau_1, \dots, \tau_{k-1}) \in \{0, 1, \dots, \tau\}^{k-1}$ , and consider the number of possible  $\vec{w} = (w^1, \dots, w^k) \in \prod_{i=1}^k \mathcal{D}_{n_i}$  such that both  $\Phi_0(\vec{w}) = v$  and  $\vec{\ell}(\vec{w}) = \vec{\tau}$ . Note that  $v$  determines the first  $N$  symbols of  $w^1 u^1 w^2 \dots u^{k-1} w^k$ , and thus it determines each symbol  $w_j^i$  which has not been truncated from the end of  $w^1 u^1 w^2 \dots u^{k-1} w^k$ . This leaves  $\sum_{i=1}^{k-1} \tau_i \leq k\tau$  remaining symbols  $w_j^i$  which are not determined by  $v$ , and thus

$$\# \left\{ w \in \prod_{i=1}^k \mathcal{D}_{n_i} \mid \Phi(\vec{w}) = v \text{ and } \vec{\ell}(\vec{w}) = \vec{\tau} \right\} \leq p^{k\tau},$$

where  $p$  is the size of the alphabet. There are at most  $(\tau + 1)^k$  choices for  $\vec{\tau}$ , and thus

$$\#\Phi_0^{-1}(v) \leq p^{k\tau}(\tau + 1)^k = (p^\tau(\tau + 1))^k,$$

which completes the proof of the lemma.  $\square$

**4.3. Proof of Proposition 3.3.** We establish Proposition 3.3, which is crucial for our lower bounds. This is the most involved stage of our proof.

**Step 0: Definition and basic properties of  $X_n$ .** First, we define the sequence of shift spaces  $X_n$  which will meet our requirements. Let  $\mathcal{G}_{\leq n} = \bigcup_{i=0}^n \mathcal{G}_i$ , and consider the set of words

$$(4.4) \quad \Phi(\mathcal{G}_{\leq n}^*) := \bigcup_{m=1}^{\infty} \{\Phi(w_1, \dots, w_m) \mid w_i \in \mathcal{G}_{\leq n}, 1 \leq i \leq m\}.$$

We can turn this set into the language of a shift space by including all subwords to obtain

$$(4.5) \quad \mathcal{L}(X_n) := \{\text{all subwords of elements of } \Phi(\mathcal{G}_{\leq n}^*)\}.$$

Then,  $X_n$  is defined as the shift whose language is  $\mathcal{L}(X_n)$  – that is,  $X_n$  is the set of sequences for which every finite subword is in  $\mathcal{L}(X_n)$ .

**Lemma 4.4.**  *$X_n$  is a well-defined shift space. Furthermore,  $X_n$  has the following properties.*

- (1) *For every  $w \in \mathcal{L}(X_n)$ , there exist  $u, v \in \mathcal{L}$  such that  $|u|, |v| \leq n + \tau$  and  $uwv \in \mathcal{G}$ , where  $\tau$  is the transition time in the  $(gcW)$ -specification property of  $\mathcal{G}$ .*
- (2)  *$X_n$  has  $(W)$ -specification.*

*Proof.* By [24, Proposition 1.3.4], to check that a collection of words is a language for some shift space, we need only check the following two conditions.

- Every subblock of  $\mathcal{L}(X_n)$  is in  $\mathcal{L}(X_n)$
- For every  $w \in \mathcal{L}(X_n)$ , there are non-empty  $u, v \in \mathcal{L}(X_n)$  so that  $uwv \in \mathcal{L}(X_n)$ .

The first condition is satisfied by definition, so we verify the second. Let  $w \in \mathcal{L}(X_n)$ . Then  $w$  is a subword of some word of the form  $\Phi(w_1, \dots, w_m)$ , and for any  $u \in \mathcal{G}_{\leq n}$ , the word  $\Phi(u, w_1, \dots, w_m, u)$  satisfies the required condition.

To check the first property claimed for  $X_n$ , we observe that every  $w \in \mathcal{L}(X_n)$  is a subword of  $w^1 v^1 \dots w^m$  for some  $w^i \in \mathcal{G}$  and  $v^i$  such that the conditions in Definition 2.3 hold. In particular, by appending

at most  $n + \tau$  symbols to either end of  $w$ , we obtain a word of the form  $w^i v^i \dots w^j \in \mathcal{G}$ .

This observation also implies that  $X_n$  has (W)-specification. Indeed, given any collection of words  $w^1, \dots, w^m \in \mathcal{L}(X_n)$ , we extend these to words  $\bar{w}^1, \dots, \bar{w}^m \in \mathcal{G}$  by appending at most  $n + \tau$  symbols to either end of the word  $w^j$ , and then use (W)-specification for  $\mathcal{G}$  to conclude that  $X_n$  has (W)-specification with transition time  $3\tau + 2n$ .  $\square$

We also include the following lemma for completeness, although it plays no role in our proof.

**Lemma 4.5.** *Suppose that we also have the following condition (which is satisfied for all our examples): For every  $v \in \mathcal{L}$ , there exist words  $u, w \in \mathcal{L}$  so that  $uvw \in \mathcal{G}$ . Then the shift spaces provided by Lemma 4.4 satisfy  $X = \overline{\bigcup X_n}$ .*

*Proof.* It suffices to show that  $\mathcal{L}(X) = \bigcup_n \mathcal{L}(X_n)$ . This is easy, since by assumption for any  $w \in \mathcal{L}$ , there are  $u, v$  such that  $uvw \in \mathcal{G}$ . Thus  $uvw \in X_{|uvw|}$ . Thus  $w \in \mathcal{L}(X_{|uvw|})$ .  $\square$

The rest of the proof of Proposition 3.3 is an extension of the approach used by Pfister and Sullivan in [32]:

- (1) Construct a subshift  $Y \subset X_n$  for some  $n \geq 1$  such that every  $\nu \in \mathcal{M}_\sigma(Y)$  is weak\*-close to  $\mu$ .
- (2) Use edit approachability of  $\mathcal{L}$  by  $\mathcal{G}$  to explicitly build a subshift  $H \subset Y$  with a rich structure.
- (3) Show that  $H$  (and hence  $Y$ ) has entropy close to  $h(\mu)$  by using this structure.
- (4) Obtain the measures  $\mu'$  as maximal entropy measures for  $Y$ .

In preparation for the above steps, fix  $\eta > 0$  and use the ergodic decomposition of  $\mu$  together with affinity of the entropy map to find  $\lambda = \sum_{i=1}^p a_i \mu_i$  such that

- the  $\mu_i$  are ergodic;
- the  $a_i$  are rational numbers in  $[0, 1]$  such that  $\sum_{i=1}^p a_i = 1$ ;
- $D(\mu, \lambda) \leq \eta$ ;
- $h(\lambda) > h(\mu) - \eta$ .

Let  $h_i = 0$  when  $h(\mu_i) = 0$ , and  $\max(0, h(\mu_i) - \eta) < h_i < h(\mu_i)$  otherwise.

**Definition 4.6.** Given  $\nu \in \mathcal{M}(X)$  and  $\zeta > 0$ , let

$$\mathcal{L}^{\nu, \zeta} := \{w \in \mathcal{L} \mid D(\mathcal{E}_{|w|}(x), \nu) < \zeta \text{ for all } x \in [w]\}.$$

Combining [31, Propositions 2.1 and 4.1], we have the following lemma.

**Lemma 4.7.** [31, Propositions 2.1 and 4.1] *There exists  $N \in \mathbb{N}$  such that for  $n \geq N$  and  $1 \leq i \leq p$ , we have  $\#\mathcal{L}_n^{\mu_i, \eta} \geq e^{nh_i}$ .*

Because  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , there is a mistake function  $g$  such that every  $w \in \mathcal{L}$  has  $v \in \mathcal{G}$  with  $\hat{d}(v, w) \leq g(|w|)$ . By Lemma 2.8, we can choose  $N$  large enough so that, in addition to the cardinality estimates in Lemma 4.7, we have the following property.

- If  $n \geq N$  and  $x, y \in X$  are such that  $\hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq \tau + g(n)$ , then  $D(\mathcal{E}_n(x), \mathcal{E}_m(y)) \leq \eta$ .

Without loss of generality, assume that  $a_i < 1$  for each  $i$ . Choose  $n$  such that we have  $n_i := a_i n \in \mathbb{N}$ ,  $n_i + g(n_i) \leq n$ , and  $n_i \geq N$  for every  $i$ , and moreover

$$(4.6) \quad \frac{n}{n + p\tau + \sum_{i=1}^p g(n_i)}(h(\lambda) - \eta) \geq h(\lambda) - 2\eta.$$

To prove the proposition, we will follow the steps listed above to show that there exists  $\mu' \in \mathcal{M}_\sigma^e(X_n)$  such that  $D(\mu, \mu') \leq 6\eta$  and  $h(\mu') > h(\mu) - 4\eta$ .

**Step 1: Definition of  $Y \subset X_n$ .** Fix  $K \in \mathbb{N}$  such that  $4/K \leq \eta$ . Now let

$$(4.7) \quad Y := \{x \in X_n \mid x_t x_{t+1} \cdots x_{t+Kn-1} \in \mathcal{L}_{Kn}^{\mu, 5\eta} \text{ for all } t \geq 0\}.$$

Then  $Y \subset X_n$  is compact and  $\sigma$ -invariant. Moreover, the following holds.

**Lemma 4.8.** *We have  $D(\mu, \nu) \leq 6\eta$  for any  $\nu \in \mathcal{M}_\sigma^e(Y)$ .*

*Proof.* Since  $\nu$  is ergodic, there exists a generic point  $x \in Y$ , that is,  $\mathcal{E}_m(x)$  converges to  $\nu$ . We choose  $L$  so that  $nK/L \leq \eta$  holds, take an arbitrary integer  $m \geq L$  and choose integers  $s$  and  $0 \leq q < Kn$  so that  $m = sKn + q$  holds. Then, using (4.7) and the inequalities  $\frac{q}{m} \leq \frac{Kn}{L} \leq \eta$ , we have

$$\begin{aligned} D(\mathcal{E}_m(x), \mu) &\leq \sum_{i=0}^{s-1} \frac{Kn}{m} D(\mathcal{E}_{Kn}(\sigma^{iKn} x), \mu) + \frac{q}{m} D(\mathcal{E}_q(\sigma^{sKn} x), \mu) \\ &\leq 5\eta + \eta = 6\eta. \end{aligned}$$

Thus taking  $m \rightarrow \infty$ , we have the lemma.  $\square$

**Step 2: Construction of  $H$ .** For brevity of notation we write  $\mathcal{D}^i = \mathcal{L}_{n_i}^{\mu_i, \eta}$ . Extend the definitions of  $n_i, \mathcal{D}^i, \mu_i, a_i$  to indices  $i > p$  by repeating periodically: that is, if  $i = pq + r$ ,  $1 \leq r \leq p$ , then  $n_i = n_r$ ,  $\mathcal{D}^i = \mathcal{D}^r$ ,  $\mu_i = \mu_r$  and  $a_i = a_r$ .

By the assumption that  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , we can define a map  $\phi_{\mathcal{G}}: \mathcal{L} \rightarrow \mathcal{G}$  such that  $\hat{d}(w, \phi_{\mathcal{G}}(w)) \leq g(|w|)$ . We can define a map  $\Phi_{EG}$  from the set of finite collections of words  $\mathcal{L}^*$  to  $\mathcal{G}$  by ‘editing then gluing’. That is, given  $(w^1, \dots, w^n) \in \mathcal{L}$ , we put  $\Phi_{EG}(w^1, \dots, w^n) = \Phi(\phi_{\mathcal{G}}(w^1), \dots, \phi_{\mathcal{G}}(w^n))$ , where  $\Phi$  is the gluing map (4.3). The map  $\Phi_{EG}$  extends to subsets of  $\mathcal{L}^{\mathbb{N}}$  in the natural way, and we consider it here with the following domain:

$$(4.8) \quad \Phi_{EG}: \prod_{j=1}^{\infty} \mathcal{D}^j \rightarrow X.$$

In other words, given  $\mathbf{w} = \{w^j\} \in \prod_{j=1}^{\infty} \mathcal{D}^j$ , let  $v^j = \phi_{\mathcal{G}}(w^j) \in \mathcal{G}$  and  $\Phi_{EG}(\mathbf{w}) = v^1 u^1 v^2 u^2 \dots$ , where  $u^j$  with  $|u^j| \leq \tau$  are the “gluing words” provided by (W)-specification. Let  $H = \phi(\prod_{j=1}^{\infty} \mathcal{D}^j)$ . Then we have  $H \subset X_n$  since  $n_i + g(n_i) \leq n$ .

A sort of periodicity is built into the definition of the sequences  $\phi(\mathbf{w})$ : the word  $v^i$  is an approximation of a suitable generic point for the measure  $\mu_i$ , and the measures  $\mu_i$  repeat periodically ( $\mu_{i+p} = \mu_i$ ). The following lemma states that following  $\phi(\mathbf{w})$  for a single “cycle” of this periodic behaviour gives a good approximation to  $\mu$ . We write  $\ell_j = \ell_j(\mathbf{w}) = |v^j u^j|$  for the length of the words associated to the index  $j$ , and observe that  $|\ell_j - n_j| \leq \tau + g(n_j)$ .

**Lemma 4.9.** *Fix  $\mathbf{w} \in \prod_{j=1}^{\infty} \mathcal{D}^j$ . For  $q \geq 0$ , let  $c_q = c_q(\mathbf{w}) = \sum_{r=1}^p \ell_{qp+r}$  be the length of the  $q$ th “cycle” in  $\phi(\mathbf{w})$  and let  $b_m = b_m(\mathbf{w}) = \sum_{q=0}^{m-1} c_q$ . Then we have  $D(\mathcal{E}_{c_m}(\sigma^{b_m} \phi(\mathbf{w})), \mu) \leq 3\eta$ .*

*Proof.* Choose  $x^j \in [w^j]$  for each  $j \in \mathbb{N}$ , so that by the definition of  $\mathcal{D}^j$ , we have  $D(\mathcal{E}_{n_j}(x^j), \mu_j) \leq \eta$ . Let  $y = \sigma^{b_m} \phi(\mathbf{w})$  and let  $d_j = \sum_{i=0}^{j-1} \ell_{mp+i}$  for  $1 \leq j \leq p$ . By the definition of  $\phi$  and the property following Lemma 4.7, we have

$$D(\mathcal{E}_{\ell_j}(\sigma^{d_j} y), \mu_j) \leq D(\mathcal{E}_{\ell_j}(\sigma^{d_j} y), \mathcal{E}_{n_j}(x^{mp+j})) + D(\mathcal{E}_{n_j}(x^{mp+j}), \mu_j) \leq 2\eta.$$

Observe that  $c_q \approx n$ : more precisely, we have

$$(4.9) \quad |c_q - n| \leq \sum_{r=1}^p |\ell_{qp+r} - n_r| \leq p(\tau + g(n)).$$



Taking convex combinations gives

$$\begin{aligned} D(\mu, \mathcal{E}_{c_m}(y)) &\leq D\left(\sum_{j=1}^p a_j \mu_j, \sum_{j=1}^p \frac{\ell_j}{c_m} \mathcal{E}_{\ell_j}(\sigma^{d_j} y)\right) + \eta \\ &\leq \left(\sum_{j=1}^p \left|a_j - \frac{\ell_j}{c_m}\right|\right) + 2\eta \leq 3\eta, \end{aligned}$$

provided  $N$  is chosen large enough such that  $n \geq n_j \geq N$ , and such that (4.9) guarantees we have  $\sum_{j=1}^p |a_j - \frac{\ell_j}{c_m}| \leq \eta$ .  $\square$

We are now in a position to show that  $H \subset Y$ . Given  $y = \phi(\mathbf{w}) \in H$  and  $t \in \mathbb{N}$ , we can choose  $m_1, m_2$  such that

$$b_{m_1-1} \leq t < b_{m_1} < b_{m_2} \leq t + Kn < b_{m_2+1},$$

and so

$$\mathcal{E}_{Kn}(\sigma^t y) = \left(\sum_{q=m_1}^{m_2-1} \frac{c_q}{Kn} \mathcal{E}_{c_q}(\sigma^{b_q} y)\right) + \xi_1 \mathcal{E}_{b_{m_1}-t}(\sigma^t y) + \xi_2 \mathcal{E}_{t+Kn-b_{m_2}}(\sigma^{b_{m_2}} y),$$

where  $0 \leq \xi_1, \xi_2 \leq \frac{n+p(g(n)+\tau)}{Kn} \leq \eta$ . Each of the empirical measures in the large sum is within  $3\eta$  of  $\mu$ , by Lemma 4.9, and thus we have

$$D(\mathcal{E}_{Kn}(\sigma^t y), \mu) \leq 5\eta.$$

In particular, this shows that  $y \in Y$ .

**Step 3. Estimation of entropy of  $H$ .** Now we use the definition of  $H$  to estimate its topological entropy. Our key tool will be the estimates obtained in Lemmas 2.6 and 4.3.

**Lemma 4.10.** *The topological entropy of  $H$  is at least  $h(\mu) - 4\eta$ .*

*Proof.* Fix  $m \in \mathbb{N}$  and set  $b' = n + p\tau + \sum_{j=1}^p g(n_j)$ . Note that

$$(4.10) \quad mb' \geq \sup_{\mathbf{w} \in \prod_{j=1}^{\infty} \mathcal{D}^j} b_m(\mathbf{w}) \text{ and } \frac{n}{b'}(h(\lambda) - \eta) \geq h(\lambda) - 2\eta$$

holds, where  $b_m(\mathbf{w})$  is as in Lemma 4.9. Moreover, since  $n = \sum_{j=1}^p n_j$  and each  $n_j \geq N$ , we have  $b' \geq pN$ .

Let  $\phi_m: \prod_{j=1}^{mp} \mathcal{D}^j \rightarrow \mathcal{L}_{mb'}$  be the map that takes  $w^1, \dots, w^{mp}$  to the first  $mb'$  symbols of  $\Phi_{EG}(\mathbf{w})$ , where  $\mathbf{w} \in \prod_{j=1}^{mp} \mathcal{D}^j$  and  $\Phi_{EG}$  is the ‘edit and glue’ map from Step 2. Note that  $\phi_m(\prod_{j=1}^{mp} \mathcal{D}^j) \subset \mathcal{L}_{mb'}(H)$ .

Now in order to estimate the entropy of  $H$ , we will use our estimates on the cardinality of  $\mathcal{D}^j$  together with a bound on  $\#\phi_m^{-1}(v)$  for  $v \in \mathcal{L}_{mb'}$ . Recall that  $\phi_{\mathcal{G}}: \mathcal{L} \rightarrow \mathcal{G}$  is a map which satisfies  $\hat{d}(w, \phi_{\mathcal{G}}(w)) \leq$

$g(|w|)$ . First we use Lemma 2.6, recalling that  $\frac{g(n)}{n} \rightarrow 0$ , to fix  $N_0$  sufficiently large so that

$$(4.11) \quad \#\{w \in \mathcal{L} \mid \phi_{\mathcal{G}}(w) = v\} \leq e^{\eta|v|/3}$$

for every  $v \in \mathcal{G}$  with  $|v| \geq N_0$ .

Now as  $w$  ranges over  $\mathcal{D}^j$ , the word  $\phi_{\mathcal{G}}(w)$  may vary in length; however, since its  $\hat{d}$ -distance from  $w$  is at most  $g(|w|)$ , the number of different lengths it can take is at most  $2g(|w|)$ . Similarly for  $\vec{w} = (w^1, \dots, w^{mp})$ : given  $\vec{w} \in \prod_{j=1}^{mp} \mathcal{D}^j$ , we have  $\Phi_{EG}(\vec{w}) \in \prod_{j=1}^{mp} \mathcal{G}_{n'_j}$  for some  $\vec{n}' = (n'_1, \dots, n'_{mp})$ . We see that each  $n'_j$  can take at most  $2g(|w^j|) = 2g(n_j) \leq 2g(n)$  different values, and so in particular, as  $\vec{w}$  ranges over  $\prod_{j=1}^{mp} \mathcal{D}^j$ , the number of different values taken by  $\vec{n}$  is bounded above by

$$(4.12) \quad (2g(n))^{mp} = e^{mp \log(2g(n))} \leq e^{mb' \frac{\log(2g(n))}{\min_j n_j}} \leq e^{\eta mb'/3},$$

where the last inequality follows from observing that  $n_j \approx a_j n$  and choosing  $N$  sufficiently large (since each  $n_j \geq N$ ).

For each choice of  $\vec{n}'$ , we may bound the multiplicity of the truncated gluing map  $\Phi_0: \prod_{j=1}^{mp} \mathcal{G}_{n'_j} \rightarrow \mathcal{L}_{\sum n'_j}$  using Lemma 4.3. Let  $\Phi_{mb'}$  be the map that truncates to the first  $mb'$  symbols (instead of  $\sum n'_j$ ). Because  $n'_j \leq n_j + g(n_j)$ , we have

$$\sum_{j=1}^{mp} n'_j \leq \sum_{j=1}^{mp} n_j + g(n_j) \leq mn + m \sum_{j=1}^p g(n_j) \leq mb'.$$

In particular, the first  $mb'$  symbols determine the first  $\sum n'_j$  symbols, and we conclude that for every  $v \in \mathcal{L}_{mb'}(H)$  and fixed choice of  $\vec{n}'$ , we have

$$\#\Phi_{mb'}^{-1}(v) \leq C^{mp} \leq C^{mb'/N}.$$

Assume that  $N$  was chosen large enough such that this bound gives  $\#\Phi_{mb'}^{-1}(v) \leq e^{\eta mb'/3}$ . Together with the previous bound, we see that for every  $v \in \mathcal{L}_{mb'}(H)$ , we have

$$\#\phi_m^{-1}(v) \leq e^{\eta mb'},$$

and thus we obtain the estimate

$$\#\mathcal{L}_{mb'}(H) \geq e^{-\eta mb'} \prod_{j=1}^{mp} (\#\mathcal{D}^j).$$

Using Lemma 4.7, it follows that

$$\begin{aligned} h(H) &\geq \left( \lim_{m \rightarrow \infty} \frac{1}{mb'} \sum_{j=1}^{mp} \log \# \mathcal{D}^j \right) - \eta \\ &\geq \left( \lim_{m \rightarrow \infty} \frac{1}{mb'} \sum_{j=1}^{mp} n_j h_j \right) - \eta \geq h(\lambda) - 3\eta \geq h(\mu) - 4\eta. \quad \square \end{aligned}$$

**Step 4: End of the proof of Proposition 3.3.** Let  $\mu'$  be an ergodic measure of maximal entropy for  $Y$ . Lemma 4.8 shows that  $D(\mu', \mu) \leq 6\eta$ , and Lemma 4.10 shows that  $h(\mu') \geq h(\mu) - 4\eta$ . Since  $Y \subset X_n$  by definition, this completes the proof of Proposition 3.3.

**4.4. Lower bounds.** Now we complete the proof of Theorem A by showing that the lower bound

$$(4.13) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X : \mathcal{E}_n(x) \in U\}) \geq \sup_{\mu \in U} q^\varphi(\mu)$$

holds for any open set  $U \subset \mathcal{M}(X)$ , where  $q^\varphi(\mu)$  is as in (1.1).

To show (4.13), it is sufficient to show that for any  $\mu \in \mathcal{M}(X)$  and any open neighborhood  $U \subset \mathcal{M}(X)$  of  $\mu$ ,

$$(4.14) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X : \mathcal{E}_n(x) \in U\}) \geq q^\varphi(\mu).$$

If  $\mu$  is not  $\sigma$ -invariant, then  $q^\varphi(\mu) = -\infty$  and so the equation (4.14) is trivial. Thus, we will prove the equation (4.14) for  $\mu \in \mathcal{M}_\sigma(X)$ .

Let  $\mu \in \mathcal{M}_\sigma(X)$  and  $\eta > 0$ . Then by Proposition 3.3, there exists an ergodic measure  $\nu \in U \cap \mathcal{M}_\sigma^e(X_k)$  for some  $k$  such that  $h(\nu) > h(\mu) - \eta$  and  $\int \varphi d\nu > \int \varphi d\mu - \eta$ . We use  $\nu$  to build a subset of  $\mathcal{E}_n^{-1}(U)$ , as follows.

Take  $\zeta > 0$  so small that  $\mathbb{B}(\nu, 2\zeta) \subset U$  and every measure  $\nu'$  in this neighbourhood has  $|\int \varphi d\nu' - \int \varphi d\nu| \leq \eta$ . In particular, for every  $w \in \mathcal{L}^{\nu, \zeta}$ , we have  $[w] \subset \mathcal{E}_n^{-1}(U)$ . Then, again by [31, Propositions 2.1 and 4.1], for all sufficiently large  $n$  we have

$$(4.15) \quad \#(\mathcal{L}_n^{\nu, \zeta} \cap \mathcal{L}(X_k)) \geq e^{n(h(\mu) - \eta)}.$$

We note that by the Gibbs property (3.1), we have

$$m[w] \geq K_k e^{-nP(\varphi) + S_n \varphi(x)}$$

for all  $w \in \mathcal{L}(X_k)_n$  and  $x \in [w]$ . In particular, when  $w \in \mathcal{L}^{\nu, \zeta}$  this yields

$$m[w] \geq K_k e^{-nP(\varphi) + n \int \varphi d\nu - n\eta} \geq K_k e^{n(-P(\varphi) + \int \varphi d\mu - 2\eta)}.$$

Using the estimate (4.15) and the fact that  $[w] \subset \mathcal{E}_n^{-1}(U)$  for every  $w \in \mathcal{L}^{\nu, \zeta}$ , we obtain

$$m(\mathcal{E}_n^{-1}(U)) \geq K_k e^{n(h(\mu) - P(\varphi) + \int \varphi d\mu - 3\eta)}.$$

Since  $\eta > 0$  was arbitrary, this establishes the lower bound (4.14).

## 5. APPLICATIONS

**5.1.  $S$ -gap shifts.** To check the conditions of Theorem B for a potential  $\varphi$  with the Bowen property, we verify the specification properties [A.1] and (I) on  $\mathcal{G}$  and  $\mathcal{G}^M$ , the edit approachability property [A.2], and the estimate (III) on  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi)$ .

**5.1.1. Specification properties.** Condition [A.1] is immediate because  $\mathcal{G}$  has (0)-specification. Condition (I) holds because a word in  $\mathcal{G}^M$  has the form  $0^n 10^{n_1} 10^{n_2} 1 \dots 10^{n_k} 10^m$ , where  $n_i \in S$  for all  $1 \leq i \leq k$ , and  $n, m \leq M$ . Thus, any word in  $\mathcal{G}^M$  can be extended to a word in  $\mathcal{G}$  by adding a uniformly bounded number of symbols at each end (the number of symbols to be added depends on  $M$ , but not on the length of the word), and this implies that  $\mathcal{G}^M$  has the (W)-specification property.

**5.1.2. Edit approachability.** Because  $S$  is infinite, we can choose for every  $n \in \mathbb{N}$  some  $s_n \in S$  such that  $\frac{s_n}{n} \rightarrow 0$  and  $s_n \rightarrow \infty$ . (Note that the same element of  $S$  may appear as  $s_n$  for multiple values of  $n$ .) Now define  $g: \mathbb{N} \rightarrow \mathbb{N}$  by  $g(n) := 2(\lceil n/s_n \rceil + s_n)$ , and observe that  $g$  is a mistake function.

Let  $z \in \mathcal{L}(X)_n$  and write  $s = s_n$ . The word  $z$  has the form

$$z = 0^k 10^{n_1} 10^{n_2} \dots 0^{n_i} 10^\ell.$$

We now change  $\leq k/s$  of the symbols  $0^k$  to form the word  $z^p := 0^i 10^s 10^s \dots 0^s 10^s$  ( $0 \leq i \leq s$ ). We also change  $\leq \ell/s$  of the symbols  $0^\ell$  to form the word  $z^s := 0^s 10^s 10^s \dots 0^s 10^j$  ( $0 \leq j \leq s$ ). We set  $w := z^p 10^{n_1} 10^{n_2} \dots 0^{n_i} 1 z^s$ ,  $u := 0^{s-i}$  and  $v := 0^{s-j} 1$ . Then we have  $\hat{d}(z, uvw) \leq 2(\lceil n/s_n \rceil + s_n) = g(n)$ , and  $uvw \in \mathcal{G}$  by the definition of  $\mathcal{G}$ . This shows that  $\mathcal{G}$  satisfies [A.2].

**5.1.3. Estimating  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi)$ .** Now we show that if  $\varphi$  is any potential with the Bowen property on an  $S$ -gap shift, then  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$ , verifying Condition (III). It is easy to see that  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ , so it suffices to show that

$$(5.1) \quad P(\varphi) > \overline{\lim}_{n \rightarrow \infty} \sup_{x \in X} \frac{1}{n} S_n \varphi(x).$$

Our strategy is to produce a large number of admissible words that are close (in the edit metric) to a given word, so that no single word can

carry full pressure. This strategy was also used to establish (5.1) for  $\beta$ -shifts in [8, Proposition 3.1]. For  $S$ -gap shifts, we must deal with a difficulty which does not occur for  $\beta$ -shifts: if  $x \in \Sigma_S$  is such that positions  $i$  and  $j$  both admit edits yielding new words  $x', x'' \in \Sigma_S$ , it may not be possible to make both edits simultaneously. This lack of independence between the possible edits means that it is more difficult to produce nearby words than in the case of  $\beta$ -shifts. Here, we state a sequence of lemmas which prove (5.1), whose proofs are given in §6.

**Lemma 5.1.** *We have  $P(\varphi) > \varphi(0)$ .*

In the following lemma, we use Lemma 5.1 to control words which have a small frequency of occurrence of the symbol 1.

**Lemma 5.2.** *There exists  $\epsilon > 0$  and a constant  $L = L(\epsilon)$  so that if  $x_1 \cdots x_n$  contains fewer than  $\epsilon n$  occurrences of the symbol 1, then  $\frac{1}{n} S_n \varphi(x) \leq \varphi(0) + L < P(\varphi) - L$ .*

We now control words which do not have a small frequency of occurrence of the symbol 1. This is where we use our strategy of creating a large number of new words by making edits. We need the following estimate, which is a consequence of Stirling's formula.

**Lemma 5.3.** *If  $\delta n \leq k \leq \frac{n}{2}$ , then  $\log \binom{n}{k} \geq -n\delta \log \delta - 2 \log n$ .*

This estimate can be used to give a lower bound on the cardinality of a set of words where we can control the Birkhoff averages of  $\varphi$ , and we can use this to estimate the pressure from below.

**Lemma 5.4.** *Given  $\epsilon$  be as in Lemma 5.2, there exists  $L' > 0$  such that whenever  $n$  is sufficiently large and  $x_1 \cdots x_n \in \mathcal{L}$  contains  $m \geq \epsilon n$  occurrences of the symbol 1, we have  $\frac{1}{n} S_n \varphi(x) < P(\varphi) - L'$ .*

We conclude from Lemma 5.2 and Lemma 5.4 that

$$\overline{\lim}_{n \rightarrow \infty} \sup_{x \in X} \frac{1}{n} S_n \varphi(x) \leq \max\{P(\varphi) - L, P(\varphi) - L'\} < P(\varphi),$$

and it is easy to verify **(III)** from this together with  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ .

**5.2.  $\beta$ -shifts.** Every  $\beta$ -shift can be presented by a countable state directed labelled graph with vertices  $v_1, v_2, \dots$ . For every  $i \geq 1$ , we draw an edge from  $v_i$  to  $v_{i+1}$ , and label it with the value  $\omega_i^\beta$ . Next, whenever  $\omega_i^\beta > 0$ , for each integer from 0 to  $\omega_i^\beta - 1$ , we draw an edge from  $v_i$  to  $v_1$  labelled by that value.

The  $\beta$ -shift can be characterised as the set of sequences given by the labels of infinite paths through the directed graph which start at

$v_1$ . For our set  $\mathcal{G}$ , we take the collection of words labelling a path that begins and ends at the vertex  $v_1$ . Thus,  $\mathcal{G}$  automatically satisfies (0)-specification, and in particular, [A.1] holds.

Let  $\varphi: \Sigma^\beta \rightarrow \mathbb{R}$  be a continuous function satisfying the Bowen property. It is shown in [8, §3.1] that conditions (I)–(III) in Theorem 3.1 hold, so it only remains to check condition [A.2].

We now show that  $\mathcal{L}$  is edit approachable (in fact, Hamming approachable) by  $\mathcal{G}$  with mistake function  $g \equiv 1$ . Let  $z \in \mathcal{L}_n$ . We set  $j := \max\{1 \leq i \leq n : z_i \neq 0\}$  and define a new word  $w \in \mathcal{L}_n$  by

$$w_i = \begin{cases} z_i & (1 \leq i \leq n, i \neq j); \\ z_j - 1 & (i = j). \end{cases}$$

It is easy to see that  $\hat{d}(z, w) = 1 = g(n)$  and  $w \in \mathcal{G}$ , which implies [A.2]. It follows that  $(\Sigma^\beta, \sigma)$  satisfies the level-2 large deviations principle with reference measure  $m_\varphi$ , and rate function  $q^\varphi$  given by (1.1).

## 6. PROOFS OF OTHER TECHNICAL RESULTS

*Proof of Lemma 2.6.* We obtain an upper bound on the number of words that can be obtained by making at most  $m$  edits to  $w$  as follows. We introduce an additional symbol  $e$  (for ‘edit’), and construct a new word  $w'$  of length  $n+m$  which contains exactly  $m$  of the symbols  $e$ , and so that  $w_1 = w'_1$ . Note that  $\binom{n+m}{n}$  is an upper bound on the number of such words  $w'$ . Now obtain a new word  $v \in \mathcal{L}$  from  $w'$  by performing exactly one of the following actions at each symbol  $e$ , and then deleting the  $e$ .

- (1) Change the symbol immediately before  $e$  to a different symbol.
- (2) Insert a symbol immediately before  $e$ .
- (3) Delete the symbol immediately before  $e$ .
- (4) Leave the symbol immediately before  $e$  unchanged.

Note that every word  $v$  which satisfies  $\hat{d}(v, w) \leq m$  can be produced by this procedure. At each symbol  $e$ , there are a total of  $2\#A + 2$  possible actions, so we see that

$$\#\{v \mid \hat{d}(v, w)\} \leq (2\#A + 2)^m \binom{n+m}{n}.$$

From Stirling’s formula there is a constant  $C'$  such that

$$|\log n! - (n \log n - n)| \leq C' \log n$$

for every  $n \in \mathbb{N}$ , and so when  $m \leq \delta n$  we have

$$\begin{aligned}
\log \binom{n+m}{n} &= \log(n+m)! - \log n! - \log m! \\
&\leq ((n+m) \log(n+m) - n \log n - m \log m) \\
&\quad + C'(\log(m+n) + \log m + \log n) \\
&= \left( n \log \frac{n+m}{n} + m \log \frac{n+m}{m} \right) + 3C' \log(m+n) \\
&\leq n (\log(1+\delta) + \delta \log(1+\delta^{-1})) + 3C' \log((1+\delta)n) \\
&= n ((1+\delta) \log(1+\delta) - \delta \log \delta) + 3C' \log((1+\delta)n).
\end{aligned}$$

Using the inequalities  $1+\delta \leq 2$  and  $\log(1+\delta) \leq \delta$ , we see that the left-hand side of (2.2) admits the bound

$$\begin{aligned}
&\#\{v \in \mathcal{L} \mid \hat{d}(v, w) \leq \delta n\} \\
&\leq (2\#A + 2)^{\delta n} e^{n((1+\delta) \log(1+\delta) - \delta \log \delta)} e^{3C' \log((1+\delta)n)} \\
&\leq (2\#A + 2)^{\delta n} e^{2n\delta} e^{n(-\delta \log \delta)} (1+\delta)^{3C'} n^{3C'},
\end{aligned}$$

which completes the proof.  $\square$

*Proof of Lemma 2.8.* Let  $\hat{g}(n) = g(n) + 1$ , so that  $\hat{g}$  is also a mistake function. Take  $x, y$  and  $m, n$  as in the hypothesis of the lemma, and let  $k = \hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq g(n)$ .

Following the set-up of the proof of the previous lemma, we obtain a new word  $w'$  by inserting the symbol  $e$  into  $k$  positions of  $x_1 \cdots x_n$  to mark where an insertion, deletion or substitution will take place to obtain  $y_1 \cdots y_m$ . We write  $w = w^1 w^2 \cdots w^{k+1}$  so that the last symbol of each  $w^i$  with  $1 \leq i \leq k$  is  $e$  (note that  $w^{k+1}$  may be the empty word). Let  $w_r^i$  be the word obtained by omitting the last two symbols from  $w^i$ , and form the word  $w_r = w_r^1 w_r^2 \cdots w_r^{k+1}$  (where  $r$  stands for ‘reduced’, and if  $|w^i| \leq 2$ , then  $w_r^i$  is the empty word). For  $n \geq 0$ , let

$$\begin{aligned}
V(n) &= \sup\{|S_{m'}\varphi(x) - S_m\varphi(y)| \mid x_1 \cdots x_n = y_1 \cdots y_n \\
&\quad \text{and } m, m' \in \{n, n+1, n+2\}\}.
\end{aligned}$$

Note that continuity of  $\varphi$  implies that  $\frac{1}{n}V(n) \rightarrow 0$ . In particular, for  $z \geq 1$ , we may write  $\epsilon(z) = \sup_{m \geq z} \frac{1}{m}V(m)$  and obtain  $\epsilon(z) \rightarrow 0$ . We will use this fact for “long” words, while for “short” words we will use the bound  $V(n) \leq 2(n+2)\|\varphi\| \leq 4(n+1)\|\varphi\|$ .



Both  $x_1 \cdots x_n$  and  $y_1 \cdots y_m$  can be obtained from  $w_r$  by inserting at most two symbols at the end of each subword  $w_r^i$ , and so

$$(6.1) \quad |S_n \varphi(x) - S_m \varphi(y)| \leq \sum_{j=1}^{k+1} V(n_j),$$

where  $n_j = |w_r^j|$ . To bound this sum, we let  $C_n = \sqrt{\frac{n}{\hat{g}(n)}}$  and break the sum into two parts, corresponding to  $n_j < C_n$  and  $n_j \geq C_n$ . We have

$$(6.2) \quad \begin{aligned} |S_n \varphi(x) - S_m \varphi(y)| &\leq \sum_{n_j < C_n} V(n_j) + \sum_{n_j \geq C_n} V(n_j) \\ &\leq \sum_{n_j < C_n} 4(n_j + 1) \|\varphi\| + \sum_{n_j \geq C_n} n_j \epsilon(C_n) \\ &\leq 4C_n \|\varphi\| \hat{g}(n) + n \epsilon(C_n), \end{aligned}$$

where the last inequality uses the fact that there are  $k + 1 \leq \hat{g}(n)$  values of  $j$  in total, and that  $\sum n_j \leq n$ .

Now we can estimate the difference in Lemma 2.8 as

$$\begin{aligned} \left| \frac{1}{n} S_n \varphi(x) - \frac{1}{m} S_m \varphi(y) \right| &\leq \frac{1}{n} |S_n \varphi(x) - S_m \varphi(y)| + \left| \frac{1}{n} - \frac{1}{m} \right| |S_m \varphi(y)| \\ &\leq 4 \|\varphi\| C_n \frac{\hat{g}(n)}{n} + \epsilon(C_n) + \frac{|m - n|}{n} \frac{1}{m} |S_m \varphi(y)| \\ &\leq 4 \|\varphi\| \sqrt{\frac{\hat{g}(n)}{n}} + \epsilon(C_n) + \frac{\hat{g}(n)}{n} \|\varphi\|. \end{aligned}$$

Because  $\hat{g}$  is a mistake function, the first and third terms go to 0 as  $n \rightarrow \infty$ , while  $C_n \rightarrow \infty$  and so the second term goes to 0 as well. This completes the proof of Lemma 2.8.  $\square$

*Proof of Proposition 2.9.* Clearly  $P(\mathcal{G}, \varphi) \leq P(\varphi)$ , so it suffices to prove the other inequality. We compare  $\Lambda_n(\mathcal{L}, \varphi)$  and  $\Lambda_n(\mathcal{G}, \varphi)$  using Lemmas 2.6 and 2.8. By edit approachability, for each  $w \in \mathcal{L}_n$  there exists  $v = v(w) \in \mathcal{G}$  such that  $\hat{d}(v, w) \leq g(|w|)$ . Lemma 2.6 tells us that given  $v \in \mathcal{G}$ , the number of words  $w \in \mathcal{L}_n$  for which  $v = v(w)$  is at most

$$C n^C (e^{C\delta} e^{-\delta \log \delta})^n,$$

where  $\delta = g(n)/n$ . In particular, for all sufficiently large  $n$  this expression is bounded above by  $e^{\delta'_n n}$ , where  $\delta'_n \rightarrow 0$ .

It follows from Lemma 2.8 that there is  $\delta_n \rightarrow 0$  such that for every  $v, w$  as above and any  $x \in [v]$ ,  $y \in [w]$ , we have

$$|S_n \varphi(x) - S_{|w|} \varphi(y)| \leq n \delta_n.$$

Together the above estimates imply that

$$\Lambda_n(\mathcal{L}, \varphi) \leq \sum_{m=n-g(n)}^{n+g(n)} \sum_{w \in \mathcal{G}_m} e^{\delta'_n n} e^{n\delta_n + \sup_{y \in [w]} S_m \varphi(y)},$$

and so in particular there is  $m \in [n - g(n), n + g(n)]$  such that

$$\Lambda_m(\mathcal{G}, \varphi) \geq \frac{1}{2g(n)} e^{-(\delta'_n + \delta_n)n} \Lambda_n(\mathcal{L}, \varphi).$$

Since  $g(n)$  is sublinear and  $\delta_n, \delta'_n \rightarrow 0$ , this implies the result.  $\square$

*Proof of Lemma 2.12.* Items 1) and 3) can be obtained by making minor modifications to the proof of Proposition 2.2 in [7, §6.2], so we omit these arguments and prove only item 2).

Let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$  and  $\tilde{\mathcal{G}} \subset \mathcal{L}(X)$  be as in Lemma 2.12 and assume that  $\mathcal{G}$  satisfies [A.2]. Let  $g: \mathbb{N} \rightarrow \mathbb{N}$  be a mistake function as in [A.2] for  $\mathcal{G}$ . Then we define a mistake function  $\tilde{g}: \mathbb{N} \rightarrow \mathbb{N}$  by  $\tilde{g}(n) = (4r + 3)g(n + 2r) + 4r$ . Take a  $\tilde{z} \in \mathcal{L}(X)_n$ . Since  $\Psi$  is surjective, there exists  $z \in \mathcal{L}(\Sigma)_{n+2r}$  so that  $\Psi(z) = \tilde{z}$ . Since  $\mathcal{G}$  satisfies [A.2], we can find  $w \in \mathcal{G}$  so that  $\hat{d}(z, w) \leq g(n + 2r)$  holds, where we recall that  $r$  is the length of the block code. We set  $\tilde{w} = \Psi(w)$ .

Because  $\hat{d}(z, w) \leq g(n + 2r)$ , there exist an integer  $K \geq n - ((2r + 1)g(n + 2r) + 2r)$  and two increasing sequences  $m_1 < \dots < m_K$ ,  $n_1 < \dots < n_K$  so that

$$z_{m_i-r} \dots z_{m_i+r} = w_{n_i-r} \dots w_{n_i+r}$$

for each  $1 \leq i \leq K$ . Because  $\Psi$  is a block code with length  $r$ , we have  $\tilde{z}_{m_i} = \tilde{w}_{n_i}$  for  $1 \leq i \leq K$ . This implies that

$$\hat{d}(\tilde{z}, \tilde{w}) \leq (n - K) + (|w| - K) \leq 2(n - K) + ||w| - n| \leq \tilde{g}(n).$$

Thus  $\mathcal{G}$  satisfies [A.2].  $\square$

*Proof of Lemma 3.2.* We have

$$\Lambda_n(\mathcal{C}, \varphi) = \sum_{w \in \mathcal{C}_n} e^{\sup_{x \in [w]} S_n \varphi(x)} \leq e^{n(\sup \varphi)} \Lambda_n(\mathcal{C}, 0),$$

and so  $P(\mathcal{C}, \varphi) \leq h(\mathcal{C}) + \sup \varphi$ . By the variational principle and the assumption (BR), we have

$$\begin{aligned} P(X, \varphi) &\geq h(X) + \inf \varphi \\ &\geq h(\mathcal{C}) + \sup \varphi \geq P(\mathcal{C}, \varphi), \end{aligned}$$

which proves the lemma.  $\square$

*Proof of Theorem C.* We assume the hypotheses of Theorem C. By Lemmas 2.12 and 2.13,  $\mathcal{L}(X)$  has a decomposition  $\tilde{\mathcal{C}}^p \tilde{\mathcal{G}} \tilde{\mathcal{C}}^s$  satisfying satisfies [A.1], [A.2], (I) and  $h(\tilde{\mathcal{C}}^p \cup \tilde{\mathcal{C}}^s) = 0$ . If  $\varphi: X \rightarrow \mathbb{R}$  is a continuous function which satisfies the Bowen property (II), and the condition  $\sup \varphi - \inf \varphi < h(X)$ , then by Lemma 3.2, (III) holds. Thus, the hypotheses of Theorem B are satisfied for the equilibrium measure  $\mu_\varphi$  on  $X$ , which immediately proves Theorem C.  $\square$

*Proof of Lemma 5.1.* Let  $V$  be such that  $|S_n \varphi(x) - S_n \varphi(y)| \leq V$  whenever  $x_1 \cdots x_n = y_1 \cdots y_n$ , and in particular  $S_n \varphi(x) \geq n\varphi(0) - V'$  for every  $x \in [0^{n-1}1]$ , where  $V' = V + \varphi(0) - (\inf \varphi)$ .

Choose  $k$  large (just how large will be determined later) and let  $n_1, n_2, \dots, n_k \in S$  be distinct. Let  $\pi$  be any permutation of the integers  $\{1, \dots, k\}$ , and let  $w_\pi$  be the word  $0^{n_{\pi(1)}} 1 0^{n_{\pi(2)}} 1 \cdots 0^{n_{\pi(k)}} 1$  of length  $N = \sum_{j=1}^k (n_j + 1)$ . The estimates in the previous paragraph give

$$S_N \varphi(y) \geq N\varphi(0) - kV'$$

for every  $y \in [w_\pi]$ . Now let  $\vec{\pi} = (\pi_1, \dots, \pi_m)$  be any sequence of  $m$  such permutations, and let  $v_{\vec{\pi}} = w_{\pi_1} \cdots w_{\pi_m}$ . Choosing any  $y_{\vec{\pi}} \in [v_{\vec{\pi}}]$ , we obtain the estimate

$$(6.3) \quad \Lambda_{mN}(\mathcal{L}, \varphi) \geq \sum_{\vec{\pi}} e^{S_{mN} \varphi(y_{\vec{\pi}})} \geq (k!)^m e^{mN\varphi(0) - mkV'}.$$

We have the general bound

$$(6.4) \quad \log(k!) = \sum_{j=1}^k \log j \geq \int_1^k \log t \, dt = k \log k - k - 1,$$

which yields

$$\log \Lambda_{mN}(\mathcal{L}, \varphi) \geq m(k \log k - k - 1) + mN\varphi(0) - mkV',$$

so that dividing by  $mN$  and sending  $m \rightarrow \infty$  we have

$$P(\varphi) \geq \varphi(0) + \frac{k}{N} \left( \log k - 1 - \frac{1}{k} - V' \right).$$

Taking  $k$  large gives the result.  $\square$

*Proof of Lemma 5.2.* By Lemma 5.1, there exists  $\epsilon > 0$  such that  $\varphi(0) + 2\epsilon V' < P(\varphi)$ , where  $V'$  is the constant from the proof of the previous lemma. Note that if  $x_1 \cdots x_n$  contains fewer than  $\epsilon n$  occurrences of the symbol 1, then

$$S_n \varphi(x) \leq n\varphi(0) + \epsilon n V',$$

and in particular

$$(6.5) \quad \frac{1}{n} S_n \varphi(x) \leq \varphi(0) + \epsilon V' < P(\varphi) - \epsilon V'.$$

Setting  $L = \epsilon V'$  gives the result.  $\square$

*Proof of Lemma 5.3.* We use the upper bound

$$\begin{aligned} \log(k!) &= \sum_{j=1}^k \log j \leq \int_1^{k+1} \log t \, dt = (k+1) \log(k+1) - k \\ &= (k \log k - k) + k \log \left(1 + \frac{1}{k}\right) + \log(k+1) \\ &\leq (k \log k - k) + (1 + \log(k+1)), \end{aligned}$$

which together with (6.4) gives, for all large  $n$ ,

$$\begin{aligned} \log \binom{n}{k} &= \log(n!) - \log(k!) - \log((n-k)!) \\ &\geq (n \log n - n - 1) - (k \log k - k) - (1 + \log(k+1)) \\ &\quad - ((n-k) \log(n-k) - (n-k)) - (1 + \log(n-k+1)) \\ &\geq nh \left(\frac{k}{n}\right) - 2 \log n, \end{aligned}$$

where  $h(\delta) = -\delta \log \delta - (1-\delta) \log(1-\delta)$ .  $\square$

*Proof of Lemma 5.4.* Now assume that  $x_1 \cdots x_n$  contains  $m \geq \epsilon n$  occurrences of the symbol 1. By considering a smaller collection of indices where the entry is 1 if necessary, we may assume that  $m \leq 2\epsilon n$ .

Given  $\delta > 0$  small (just how small will be determined later), let  $\delta m < k < 2\delta m$ . Let  $R$  be the set of indices in which  $x_1 \cdots x_n$  has a nonzero symbol, and let  $\mathcal{Z}$  be the collection of subsets of  $R$  with exactly  $k$  elements.

We define a map  $\phi: \mathcal{Z} \rightarrow X$  as follows. Fix  $n_1 \neq n_2 \in S$ . Given  $Z \in \mathcal{Z}$ , at each index  $k \in Z$  insert the word  $0^{n_1}1$  into  $x$ , unless  $x_{k+1} \cdots x_{k+n_1+1} = 0^{n_1}1$ , in which case insert the word  $0^{n_2}1$ . This is allowed by the definition of the  $S$ -gap shift, and we note that the map  $\phi$  is injective.

Let  $\ell = \max\{n_1, n_2\} + 1$ , and observe that  $\phi(Z)$  is obtained from  $x$  by inserting at most  $k\ell$  symbols, so that if  $p$  is the size of the alphabet, then the map  $\Phi: \mathcal{Z} \rightarrow \mathcal{L}_n$  obtained by truncating  $\phi(Z)$  to the first  $n$  symbols has the property that  $\#\Phi^{-1}(w) \leq p^{k\ell}$  for each  $w \in \mathcal{L}_n$ .

We conclude that the map  $\Phi$  yields at least  $\binom{m}{k} p^{-k\ell}$  words  $w$  in  $\mathcal{L}_n$  with the property that

$$S_n\varphi(y) \geq S_n\varphi(x) - k\ell V' \geq S_n\varphi(x) - 4\epsilon\delta nV'$$

for every  $y \in [w]$ . In particular, together with Lemma 5.3 and the conditions on  $m$  and  $k$ , this gives the estimate

$$\begin{aligned} \log \Lambda_n(\mathcal{L}, \varphi) &\geq -m\delta \log \delta - 2 \log m - k\ell \log p + S_n\varphi(x) - 4\epsilon\delta nV' \\ &\geq (\epsilon n)(-\delta \log \delta) - 4 \log(\epsilon n) - 4\epsilon\delta \ell \log p + S_n\varphi(x) - 4\epsilon\delta nV'. \end{aligned}$$

Dividing by  $n$  gives

$$\frac{1}{n} \log \Lambda_n(\mathcal{L}, \varphi) \geq \frac{1}{n} S_n\varphi(x) + \epsilon\delta(-\log \delta - 4\ell \log p - 4V') - 4 \frac{\log(\epsilon n)}{n},$$

which yields the desired result when  $\delta$  is chosen sufficiently small and  $n$  is chosen sufficiently large.  $\square$

#### APPENDIX A. FROM (S)-SPECIFICATION TO (W)-SPECIFICATION

We fill in the details of the proof that Theorem 3.1 holds as stated, with the assumption that each  $\mathcal{G}^M$  has (W)-specification. This theorem appears as Theorem C of [7] with the stronger assumption that each  $\mathcal{G}^M$  has (S)-specification.

In [7], the hypothesis of (S)-specification for  $\mathcal{G}^M$  is used in exactly three places: in the proofs of Lemma 5.1, Proposition 5.5, and Lemma 5.9. We describe how Lemma 4.3 lets us prove these intermediate results using (W)-specification for  $\mathcal{G}^M$ .

Given  $M > 0$ , let  $(\mathcal{G}^M)^*$  denote the set of finite sequences  $(w^1, \dots, w^k)$ , where each  $w^j \in \mathcal{G}^M$ . Define a map  $\Phi: (\mathcal{G}^M)^* \rightarrow \mathcal{L}$  as in (4.3), and let  $\Phi_0(\vec{w})$  be the truncation of  $\Phi(\vec{w})$  to its first  $N$  symbols, where  $N = \sum_{i=1}^k |w^i|$ . By Lemma 4.3, for every  $M > 0$  there exists a constant  $C_M > 0$  such that for each  $n_1, \dots, n_k \in \mathbb{N}$ , the map

$$\Phi_0: \prod_{i=1}^k (\mathcal{G}^M)_{n_i} \rightarrow \mathcal{L}_N, \quad N = \sum_{i=1}^k n_i$$

has the property that

$$(A.1) \quad \#\Phi_0^{-1}(v) \leq C_M^k \text{ for every } v \in \mathcal{L}_N.$$

For [8, Lemma 5.1], we can repeat the argument from that paper and apply (A.1), where  $\Phi_0$  has the domain  $\prod_{i=1}^k (\mathcal{G}^M)_n$  to get the bound

$$\Lambda_{kn}(\mathcal{L}, \varphi) \geq \Lambda_n(\mathcal{G}^M, \varphi)^k (e^{-(V_M + t_M \|\varphi\|)} C_M^{-1})^k,$$

which yields

$$\frac{1}{n}\Lambda_n(\mathcal{G}^M, \varphi) \leq P(\varphi) + \frac{1}{n}(V_M + t_M\|\varphi\| + \log C_M)$$

and shows that there exists  $D_M > 0$  with  $\Lambda_n(\mathcal{G}^M, \varphi) \leq D_M e^{nP(\varphi)}$  for all  $n$ .

For [8, Proposition 5.5], we follow the proof in that paper and take  $M$  such that  $\Lambda_n(\mathcal{G}^M, \varphi) \geq C e^{nP(\varphi)}$  for all  $n$ . Fixing  $w \in \mathcal{G}^M$ , we let  $n = |w|$  and estimate  $\nu_m(\sigma^{-k}(w))$  for  $0 \leq k \leq m$ . Let  $\ell = m - k - n$  and define maps

$$\Phi: (\mathcal{G}^M)_k \times (\mathcal{G}^M)_\ell \rightarrow \mathcal{L}_m, \quad \vec{\tau}: (\mathcal{G}^M)_k \times (\mathcal{G}^M)_\ell \rightarrow \{0, 1, \dots, t_M\}^2$$

by  $\Phi(w^1, w^2) = v_1 \dots v_m$ , where  $v = w^1 u^1 w u^2 w^2$  and  $\vec{\tau}(w^1, w^2) = (|u^1|, |u^2|)$ . Now there exist  $\tau_1, \tau_2 \in \{0, 1, \dots, t_M\}^2$  such that

$$\sum_{\substack{w^1 \in \mathcal{G}_k^M, w^2 \in \mathcal{G}_\ell^M \\ \vec{\tau}(w^1, w^2) = (\tau_1, \tau_2)}} e^{\varphi_m(\Phi(v_1, v_2))} \geq \frac{1}{(t_M + 1)^2} \sum_{w^1 \in \mathcal{G}_k^M, w^2 \in \mathcal{G}_\ell^M} e^{\varphi_m(\Phi(v_1, v_2))},$$

where  $\varphi_m(u) = \sup_{x \in [u]} S_m \varphi(x)$  for  $u \in \mathcal{L}_m$ . In particular, the same computation as in [8] gives

$$\nu_m(\sigma^{-(k+\tau_1(k))}([w])) \geq \frac{C^2 C_2^{-1}}{(t_M + 1)^2} e^{-(3V+2t_M\|\varphi\|)} e^{-2t_M P(\varphi)} e^{-nP(\varphi) + \varphi_n(w)}.$$

Summing over  $k$  gives the desired lower bound for  $\nu_m$  (at the cost of a further factor of  $(t_M + 1)^{-1}$ ), and passing to  $m \rightarrow \infty$  gives the Gibbs bound for  $\mu$ .

It remains only to prove [8, Lemma 5.9]; the proof is similar to [8, Proposition 5.5]. We show that for all sufficiently large  $M$ , there exists  $E_M > 0$  such that if  $u, v \in \mathcal{G}^M$ , then

$$(A.2) \quad \overline{\lim}_{n \rightarrow \infty} \mu([u] \cap \sigma^{-n}[v]) \geq E_M \mu(u) \mu(v).$$

In [8], Lemma 5.9 provides the inequality (A.2) with a  $\liminf$  instead of a  $\limsup$ , which is a slightly stronger result. This lemma is only used in the proof of [8, Proposition 5.8], which establishes ergodicity of the equilibrium state. One can easily verify that this ergodicity proof only requires the  $\overline{\lim}$  in the inequality (A.2).

To this end, we take  $M$  large enough that  $\Lambda_n(\mathcal{G}^M, \varphi) \geq C e^{nP(\varphi)}$ , fix  $u, v \in \mathcal{G}^M$ , take  $m \in \mathbb{N}$  large, and fix  $k \leq m$ . We estimate

$$(\nu_m \circ \sigma^{-\hat{k}})([u] \cap \sigma^{-\hat{n}}[v]) = \nu_m(\sigma^{-\hat{k}}[u] \cap \sigma^{-(\hat{n}+\hat{k})}[v])$$

for some  $\hat{k} \in [k, k + t_M]$  and some  $\hat{n} \in [n, n + 2t_M]$ . We let  $\ell = m - n - k - |u| - |v|$  and observe that (W)-specification allows us to associate to every  $(w^1, w^2, w^3) \in (\mathcal{G}^M)_k \times (\mathcal{G}^M)_n \times (\mathcal{G}^M)_\ell$  a word

$v \in \mathcal{L}_m$  such that  $v$  is the first  $m$  symbols of  $w^1 x^1 u x^2 w^2 x^3 v x^4 w^3$ , where  $|x^i| = \tau_i \in \{0, 1, \dots, t_M\}$ . As in the modified proof of [8, Proposition 5.5], we can choose  $(\tau_1, \tau_2, \tau_3, \tau_4)$  such that the set of triples  $(w^1, w^2, w^3)$  for which  $|x^i| = \tau_i$  has weight at least  $(t_M + 1)^{-4}$  of the total (over all triples), and then putting  $\hat{k} = k + \tau_1$  and  $\hat{n} = \tau_2 + n + \tau_3$ , we have

$$(\nu_m \circ \sigma^{-\hat{k}})([u] \cap \sigma^{-\hat{n}}[v]) \geq \frac{C^3 C_2^{-1}}{(t_M + 1)^4} e^{-4t_M P(\varphi)} e^{-(3V + 4t_M \|\varphi\|)} \mu(u) \mu(v).$$

Summing over  $k$  gives us the result for  $\mu_m$  at the cost of another factor of  $(t_M + 1)^{-1} (2t_M + 1)^{-1}$ , and then sending  $m \rightarrow \infty$  gives (A.2).

## REFERENCES

- [1] V. Araújo and M.J. Pacifico. Large deviations for non-uniformly expanding maps. *J. Stat. Phys.* 125 (2006), no. 2, 415–457.
- [2] R. Bowen, Periodic points and measures for Axiom A diffeomorphisms, *Trans. Amer. Math. Soc.* 154, 377–397 (1971).
- [3] H. Cramér, Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles*, 736, 5–23 (1938).
- [4] Y. M. Chung. Large deviations on Markov towers. *Nonlinearity* 24 (2011), 1229–1252.
- [5] Y. M. Chung and H. Takahasi. Large deviation principle for Benedicks-Carleson quadratic maps: *Comm. Math. Phys.* 315 (2012) 803–826.
- [6] V. Climenhaga, Topological pressure for simultaneous level sets, *Nonlinearity*, 26, (2013) 241–268.
- [7] V. Climenhaga and D. J. Thompson, Intrinsic ergodicity beyond specification:  $\beta$ -shifts,  $S$ -gap shifts, and their factors, *Israel J. of Math.* 192, no. 2, 785–817 (2012)
- [8] V. Climenhaga and D. J. Thompson, Equilibrium states beyond specification and the Bowen property, *J. London Math. Soc.* 87, no. 2 (2013) 401–427
- [9] V. Climenhaga and D. J. Thompson, Intrinsic ergodicity via obstruction entropies, to appear in *Ergodic Theory Dynam. Systems*.
- [10] H. Comman and J. Rivera-Letelier. Large deviation principles for non-uniformly hyperbolic rational maps. *Ergod. Th. & Dynam. Sys.* 31 (2011), 321–349.
- [11] D. Ahmadi Dastjerdi and S. Jangjoo, Dynamics and topology for  $S$ -gap shifts, *Topology and its Applications*, 159 (2012), 2654–2661.
- [12] M. Denker, Large deviation and the pressure function, *Transactions of the 11th Prague Conference on Information theory, Statistical Decision Functions, Prague, 1990*, Academia Publ. House of the Czechoslovak Acad. of Science, 21–33 (1992).
- [13] A. Eizenberg, Y. Kifer and B. Weiss, Large deviations for  $\mathbb{Z}^d$ -actions, *Comm. Math. Phys.* 164, 433–454 (1994).
- [14] R.S. Ellis, Large deviations for a general class of random vectors. *Ann. Probab.* 12 (1984), no. 1, 1–12.
- [15] R.S. Ellis, *Entropy, Large deviations and Statistical Mechanics*, Grundlehren der Mathematischen Wissenschaften 271, Springer-Verlag, New York, 1985.



- [16] R.S. Ellis, The Theory of Large Deviations and Applications to Statistical Mechanics, in “Long-Range Interacting Systems: Les Houches 2008 Session XC, Oxford University Press (New York), 227–277 (2010).
- [17] H. Föllmer and S. Orey. Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.* 16 (1988), no. 3, 961–977.
- [18] J. Gärtner, On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.* 22 (1977), no. 1, 27–42.
- [19] H. Hennion and L. Hervé. Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness. *Lecture Notes in Mathematics*, 1766. Springer-Verlag, Berlin, 2001.
- [20] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes études Sci. Publ. Math.* No. 51 (1980), 137–173.
- [21] G. Keller and T. Nowicki. Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. *Comm. Math. Phys.* 149 (1992), no. 1, 31–69.
- [22] Y. Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* 321, 505–524 (1990).
- [23] Y. Kifer, Large deviations, averaging and periodic orbits of dynamical systems *Comm. Math. Phys.* 162, 33–46 (1994).
- [24] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1995.
- [25] A.O. Lopes. Entropy and large deviation. *Nonlinearity* 3 (1990), no. 2, 527–546.
- [26] I. Melbourne, Large and moderate deviations for slowly mixing dynamical systems, *Proc. Amer. Math. Soc.*, 37, no.5, (2009), 1735–1741.
- [27] I. Melbourne and M. Nicol, Large deviations for nonuniformly hyperbolic systems, *Trans. Amer. Math. Soc.*, 360, no.12 (2008), 6661–6676.
- [28] K. Moriyasu, K. Sakai and K. Yamamoto, Regular maps with the specification property, *Discrete Contin. Dyn. Syst.* 33 (2013), no.7, 2991–3009.
- [29] S. Orey and S. Pelikan, Large deviation principles for stationary processes, *Ann. Probab.* 16, no. 4, 1481–1495 (1988).
- [30] S. Orey and S. Pelikan, Deviation of trajectory averages and the defect in Pesin’s formula for Anosov diffeomorphisms, *Trans. Amer. Math. Soc.* 315 (1989), 741–753.
- [31] C.-E. Pfister and W.G. Sullivan, Large Deviations Estimates for Dynamical Systems without the Specification Property. Application to the  $\beta$ -shifts, *Nonlinearity* 18 (2005), 237–261.
- [32] C.-E. Pfister, W.G. Sullivan, On the Topological Entropy of Saturated Sets, *Ergodic Theory Dynam. Systems* 27, 929–956 (2007).
- [33] M. Pollicott and R. Sharp. Large deviations for intermittent maps. *Nonlinearity* 22 (2009), no. 9, 2079–2092.
- [34] M. Pollicott, R. Sharp, and M. Yuri. Large deviations for maps with indifferent fixed points. *Nonlinearity* 11 (1998), no. 4, 1173–1184.
- [35] L. Rey-Bellet and L.-S Young, Large deviations in non-uniformly hyperbolic dynamical systems, *Ergodic Theory Dynam. Systems*, 28 (2008), 578–612.
- [36] J. Schmeling, Symbolic dynamics for  $\beta$ -shifts and self-normal numbers, *Ergodic Theory Dynam. Systems* 17, 675–694 (1997).

- [37] Y. Takahashi, Entropy functional (free energy) for dynamical systems and their random perturbations. Stochastic analysis (Katata/Kyoto, 1982), North-Holland Math. Library 32, North-Holland, Amsterdam, 437–467 (1984).
- [38] Y. Takahashi, Two aspects of large deviation theory for large time, in “Probabilistic methods in mathematical physics (Katata/Kyoto, 1985)”, Academic Press, Boston 363–384 (1987).
- [39] F. Takens, E. Verbitskiy, On the variational principle for the topological entropy of certain non-compact sets, *Ergod. Th. & Dynam. Sys.* 23, 317–348 (2003).
- [40] D.J. Thompson, Irregular sets, the  $\beta$ -transformation and the almost specification property, *Trans. Amer. Math. Soc.* 364, 5395–5414 (2012).
- [41] P. Varandas, Non-uniform specification and large deviations for weak Gibbs measures, *J. Statist. Phys.*, 146, 330–358 (2012).
- [42] P. Walters. Equilibrium states for  $\beta$ -transformations and related transformations. *Math. Z.* 159 (1978), no. 1, 65–88.
- [43] K. Yamamoto, On the weaker forms of the specification property and their applications, *Proc. Amer. Math. Soc.* **137**, no. 11, 3807–3814 (2009).
- [44] L.-S. Young, Some large deviation results for dynamical systems, *Trans. Amer. Math. Soc.* 318, no. 2, 525–543 (1990).
- [45] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* 147, 585–650 (1998).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON, HOUSTON, TEXAS 77204, USA

*E-mail address:* `climenha@math.uh.edu`

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 100 MATH TOWER, 231 WEST 18TH AVENUE, COLUMBUS, OHIO 43210, USA

*E-mail address:* `thompson@math.osu.edu`

SCHOOL OF INFORMATION ENVIRONMENT, TOKYO DENKI UNIVERSITY, 2-1200 MUZAIGAKUENDAI, INZAI-SHI, CHIBA 270-1382, JAPAN

*E-mail address:* `yamamoto.k.ak@sie.dendai.ac.jp`